DUKE
UNIVERSITY

LIBRARY

DUKE UNIVERSITY LIBRARY

MANUSCRIPT THESES

This volume may be consulted freely, but the literary rights of the author must be respected. No passage may be copied or closely paraphrased without the previous written consent of the author. If the reader obtains assistance from this volume he must give credit in his own work.

This thesis by Gary Bruce Cox has been used by the following persons, whose signatures attest their acceptance of the above restrictions.

[A library borrowing this thesis for use by one of its patrons should secure the signature of the user]

NAME	ADDRESS	DATE
[illegible]	[illegible]	4/1/81 - [illegible]

Digitized by the Internet Archive
in 2016

https://archive.org/details/cognitivestructu00coxg

Duke University Library

The use of this thesis is subject to the usual restrictions that govern the use of manuscript material. Reproduction or quotation of the text is permitted only upon written authorization from the author of the thesis and from the academic department by which it was accepted. Proper acknowledgment must be given in all printed references or quotations.

FORM 418 1M 11-43

COGNITIVE STRUCTURE: A COMPARISON OF TWO THEORIES AND MEASURES OF INTEGRATIVE COMPLEXITY

by

Gary Bruce Cox

Department of Psychology
Duke University

Date: December 19, 1969

Approved:

Michael A. Wallach

Michael A. Wallach, Supervisor

John Altrocchi

Edward E. Jones

C. R. Shuey Jr.

D. S. Burdick

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology in the Graduate School of Arts and Sciences of Duke University

1969

ABSTRACT

(Psychology-Social)

COGNITIVE STRUCTURE: A COMPARISON OF TWO THEORIES AND MEASURES OF INTEGRATIVE COMPLEXITY

by

Gary Bruce Cox

Department of Psychology
Duke University

Date: December 19, 1969

Approved:

Michael A. Wallach

Michael A. Wallach, Supervisor

John Altracchi

Edward E. Jones

C. H. Shiny Jr.

D. S. Burdick

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology in the Graduate School of Arts and Sciences of Duke University

1969

Ph.D.
C877C
1970

ABSTRACT

COGNITIVE STRUCTURE: A COMPARISON OF TWO THEORIES AND MEASURES OF INTEGRATIVE COMPLEXITY

by

Gary Bruce Cox

This study was intended to assess the generality of a particular type of cognitive structure characteristic, that of integrative complexity. Pursuant of this, the theories of H. M. Schroder and O. J. Harvey were compared, and their respective measures administered to 440 Ss of both sexes from three different southern schools.

The theoretical analysis suggested that Schroder's theory is more truly structural in nature, and is more powerful in that it is more easily and obviously applicable to a broad range of cognitive domains. Harvey's position is much more firmly grounded on the content of the interpersonal domain. Both theories claim that the characteristic of cognitive structure which is most important in determining cognitive complexity is not differ-

entiation, or an increase in the dimensionality, of the cognitive domain, but the subsequent integration of the differentiated components. Unfortunately, neither theorist is able to define integration so as to clearly distinguish it from a dimensional position. Here again, however, Schroder's theory seems to be the stronger, since it is at least explicit enough that the locus of difficulty can be precisely identified. Further, even if Schroder is unable to define adequately the integration concept, his theorizing suggests the importance of the possibility of super- and sub-ordinate relationships among dimensions.

Results of the testing were as follows: (a) As expected, the respective measures of cognitive integration were nonsignificantly correlated with each other. (b) Both measures of integration were significantly correlated with such measures of intelligence as vocabulary, abstract thinking, and SAT verbal and mathematical scores. (c) Sex differences in the scores may exist, although the pattern is not clear. (d) Other sample characteristics may affect the distribution of scores, e.g. large intelligence differences, socioeconomic differences, etc. On the other hand, Negroes are not ipso facto inferior to Caucasians, even when the latter enjoy a 100-point advantage on SAT averages. Nor are Southern whites inferior to Northern whites, at least when both are of superior intellectual ability. (e) Reliability, as estimated by coefficient alpha, is satisfactory for Harvey's measure and unsatisfactory for Schroder's. (f) The distribution of scores is such that for both measures complex Ss are rare, so pools of Ss

must be tested in order to obtain adequate numbers of complex Ss. This is more a problem in attempting to apply Schroder than Harvey, largely because Schroder has often not bothered to study middle-range Ss, so their characteristics are unknown. Schroder's variable (especially) is essentially inapplicable to an unscreened group of subjects.

Overall, Schroder's theory seems more promising than Harvey's. Suggestions were made for improving the reliability and distribution of scores. Additionally, a translation of Schroder's theory into dimensional terminology was attempted, and some important implications of his position for the dimensional orientation were discussed.

ACKNOWLEDGMENTS

I would like to thank my committee chairman, Dr. Michael Wallach, for the support and encouragement he has offered throughout my work at Duke. Dr. John Altrocchi has also consistently shown a willingness to talk, which I have appreciated, along with the efforts of the other members of my committee, Drs. Clifford Wing, Edward Jones, and Donald Burdick. A special note of thanks should go to Dr. J. R. Law, an unusually cooperative person, for the time he has made available to me.

I am also grateful to the people on whom the burden of this dissertation has fallen secondmost heavily, namely my wife and my students at Johnson C. Smith University, none of whom expressed open irritation during my periods of preoccupation and neglect.

Both H. M. Schroder and O. J. Harvey supplied me with useful materials. Dr. Schroder was especially helpful, and in addition to him I would like to thank Gary Gardiner, Jacquie Phares and Jack Carrington at Princeton for their hospitality as well as their suggestions.

In Durham, Bob and Joanna Morris frequently opened their home and refrigerator to me, and Edna Bissette did my typing, all of which I deeply appreciate.

Finally, I would like to thank Rudolph Cox, who made life bearable by occasionally insisting on a break.

G. B. C.

CONTENTS

ABSTRACT iii

ACKNOWLEDGMENTS vi

LIST OF TABLES x

I. INTRODUCTION 2

Components of Complexity 3
The Generality of Cognitive Complexity 7
Summary 14

II. CONCEPTUAL SYSTEMS: THEORY 17

The Theory of Conceptual Systems: Harvey, Hunt and Schroder 17
Conceptual Systems: Theoretical Modifications 25

III. CONCEPTUAL SYSTEMS: OPERATIONS 32

Schroder's Paragraph Completion Test (Schroder, Driver & Streufert, 1967) 32
Harvey's "This I Believe" Test (Harvey, 1966) 39
A Comparison of the Two Positions 47
The Present Study 49

IV. METHOD 51

Subjects 51
Measures and Procedure 52
Scoring 54

V. RESULTS 58

Preliminary Tests 62
Complexity, Age, College Class, Sex and Residence 66
Complexity, Schools and Intelligence 68
PCT, TIB and Intelligence 72
Reliability 77

VI. DISCUSSION 80

Summary of Results 80
Comment and Implications 85

APPENDIX 92

REFERENCES 97

LIST OF TABLES

Table

1.	Means, Standard Deviations and N's	60
2.	Complexity Score Intercorrelations and N's	61
3.	Preliminary Tests	63
4.	UNC-C Sample Homogeneity	64
5.	Sex x Schools	70
6.	Sex x Schools	71
7.	Correlations Between Intelligence Measures and the Paragraph Completion Test (PCT)	74
8.	Correlations Between Intelligence Measures and the "This I Believe" Test (TIB)	75
9.	Alpha Coefficients	79

COGNITIVE STRUCTURE: A COMPARISON OF TWO THEORIES AND MEASURES OF INTEGRATIVE COMPLEXITY

Chapter I

INTRODUCTION

In 1961 O. J. Harvey and H. M. Schroder collaborated, with D. E. Hunt, in writing the book Conceptual Systems and Personality Organization (Harvey, Hunt & Schroder, 1961). Since that time both Schroder and Harvey have actively elaborated the subject. This paper is a comparison of their theories and measures of cognitive complexity (or conceptual systems to use Harvey's term, or conceptual structure to use Schroder's).

The need for this comparison results from the common assumption that the two men are still dealing with the same theory, an impression which is strengthened by their tendencies to refer to and cite each other's work as support for their own.

Both theories are concerned with the topic of cognitive complexity or cognitive structure, a research area which is having considerable trouble deciding what it means, operationally, by its own terms. There is some question as to whether the various measures of cognitive complexity are related to each other. This concern has taken the form of a number of studies testing the "generality of cognitive complexity," that is, whether

various measures produce comparable results. The present study may be viewed as an example of such work. Accordingly, it is important to ask what evidence already exists concerning generality and how it relates to the specific question of generality at hand. As a preliminary it is necessary to understand something about the theoretical components of cognitive complexity. These will be the two functions of this chapter.

Components of Complexity

The cognitive structure theorists share the opinion that in attempting to explain and predict behavior it will be fruitful to look at the structure or pattern of, or relationships among, cognitive elements (opinions, beliefs, attitudes, ideas, etc.) in addition to the content of the elements. Beyond this basic orientation agreement diminishes. Generally, complexity and/or structure is regarded as being a function of differentiation, articulation or discrimination, and integration (Bieri, 1961; Schroder, Driver & Streufert, 1967). Other writers (e.g. Scott, 1969) have culled additional structural concepts from the literature ("centrality," etc.), but these will be considered here only very briefly.

Differentiation may be defined either in terms of dimensions (Bieri, 1955) or categories (Gardner, 1953), and in either case represents the number of stable, unique and important units (dimensions or categories) in

a particular cognitive realm (content area). Dimensional differentiation is usually measured with the Kelly Role Repertory Test (REP) (Bieri, 1961; Kelly, 1955) or with multidimensional scaling (Schroder et al., 1967). The semantic differential, with or without factor analysis, might also be used. Categorical differentiation may be assessed using object-sorting tasks, either without replacement (an object may appear in only one category--Clayton & Jackson, 1961; Gardner, 1953) or with replacement (an object may appear in as many categories as S wishes--Scott, 1962).

Unfortunately, there is no obvious relationship between the two types of differentiation, nor, for that matter, between the two ways of measuring categorical differentiation. Gardner's scoring procedure is unsatisfactory (see Messick & Kogan, 1963), and the researcher interested in object sorting would probably be better off with an adaption of Scott's (1962) technique. Since categories may be based on either different dimensions or different locations along a single dimension, some further measurement steps would be required to relate categories and dimensions.

The justification for the various dimensional measures is fairly straightforward. The REP is probably the most commonly used instrument for assessing cognitive complexity/structure and has been used in so many different ways that a separate chapter of discussion would be necessary to explore the subject. As a result, one must be chary of interpreting results based on the REP unless the way the test was used is also known.

Multidimensional scaling is related to dimensionality in that one

learns the number of dimensions necessary to support the relationships reported by the subject. Due to the differences in the models underlying scaling and factor analysis, it is not appropriate to name the dimensions for the former, although this is typically the goal in the latter.

The second property of cognitive structures is articulation (Bieri, 1961, and Schroder et al., 1967, prefer the term "discrimination"), which means the number of distinctions one makes along a particular dimension. Articulation is expected to vary from one dimension to another. Also, it is a dimensional and not a categorical characteristic.

Perhaps the best measure (in terms of its flexibility and its consistency with dimensional techniques) of articulation is that of Carr's (1965, Carr also prefers the term "discrimination"; Scott, 1969, uses a similar measure). Degree of articulation is to some extent controlled in the semantic differential and in most versions of the REP. Both Scott (1967) and Schroder et al. (1967) suggest that variables such as equivalence range (Sloane, Gorlow & Jackson, 1963) and category width (Pettigrew, 1958) may be related to articulation. Scott leaves it with that, but Schroder et al. attempt to incorporate the findings that abstract Ss have broad "category width" scores (e.g. Clayton & Jackson, 1961; Pettigrew, 1958). The theoretical explanation for this phenomenon is that, given the choice between many narrow dimensions or a few broad ones, the S with abstract schemata will ". . . lean toward greater integration--that is, greater range of stimuli per dimension--and, hence, to fewer dimensions . . ." (Schroder et al.,

1967, p. 175). No real explanation or theoretical derivation is offered as to why integration is expected, nor is there any specification of how to recognize situations in which discrimination will be evidenced by integration vs. those in which it will appear as more and finer gradations. What, for example, should we expect in the Scott (1962, 1969) sorting task which restricts neither the size nor number of groups? Also, Messick and Kogan (1963) have shown that the Clayton and Jackson type of sorting results are due more to the failure to categorize than to the number of categories formed. So far as I know, relationships between size of true categories (two or more items) with other variables, a more relevant question and one to which Scott's procedure is applicable, has not been investigated. Apropos this topic, Wallach[1] has suggested that it is important to distinguish situations in which Ss are given a category and asked to fit cases to it (as Pettigrew, 1958) from those in which S is permitted to provide his own categories or dimension (as Clayton & Jackson, 1961). More broadly, the important variable may be whether S construes the situation as one in which he is supposed to find similarities or one in which he is supposed to find differences, the complex S presumably being better at either.

The third characteristic of cognitive structures, and the last which we will consider, is integration. Although integration is mentioned by other theorists (e.g. Scott, 1969), Harvey et al. (1961) stand alone in the

---

[1] M. A. Wallach, personal communication.

degree of emphasis they put upon it as the most important determinant of complexity/abstractness. Scott's attempts (1966, 1969) to measure integration differ radically from those of either Harvey or Schroder and will not be discussed at length. (Similarly for Zajonc, 1960, who, from Scott's viewpoint may be dealing partially with integration.) The Schroder/Harvey approaches will be dealt with more extensively in Chapter III. In passing, it should be noted that the differences in emphasis on integration are not a logical consequence of theoretical differences. This is particularly true for James Bieri, who has roots in George Kelly's theory of personal constructs (Kelly, 1955), which is strikingly similar to, and in some respects more systematic than, Harvey et al. (1961).

The Generality of Cognitive Complexity

Questions about the generality of cognitive complexity have two distinguishable components, related respectively to theory and operations. Essentially all theorists agree that complexity will vary across cognitive domains or content areas (Bieri, 1961; Gardner & Schoen, 1962; Scott, 1962) or within the same content area across time (Zajonc, 1960) or as a function of situational variables (Schroder et al., 1967). Beyond this, however, is the methodological question of whether various measures of complexity at a particular time and with regard to a particular topic would

order Ss similarly.

Among measures of complexity which emphasize dimensionality and/or discrimination, some findings favoring generality have been reported. For example, Bieri and Blacker (1956) found complexity as measured by the REP test and by Rorschach perceptions to be related. Allard and Carlson (1963) administered three different REP tests, which correlated at moderately high levels. Hess (1966) compared REP complexity and performance on a sorting task and reported that the two were related. Todd and Rappoport (1964) discovered that two different models of cognitive structure (factor analysis and Hays' implication model) produced highly similar interstimulus distances which, in turn, were related to sorting performance. Since these findings are somewhat tangential to our major interests, no particular effort will be made to evaluate them.

More immediately relevant results which favor generality are reported by Carr (1965) who states that conceptual system level is related to discrimination, but not to REP fluency, and Harvey's (1966) and Wolfe's (1963) findings of relationships between REP scores (not fluency) and conceptual level.

I know of three studies of generality which are more extensive in scope than the above. Of these, Scott's (1969) is not particularly relevant to the immediate purpose here, because his theoretical orientation is dimensional, and he includes neither Schroder's nor Harvey's measure of integration. In addition to dimensionality and articulation, Scott does

measure what he defines as integrative characteristics (Centrality, Affective Balance, and Affective-Evaluative Consistency), but each of these implies a reduction in the dimensionality of the cognitive space, a position very different from that of Schroder or Harvey. Scott measured each of the five characteristics above in several different ways and found modest intra-characteristic correlations across the various scores. The characteristics correlated with each other in the predicted pattern, but this evidence for generality within Scott's model does not bear on the question of the generality of Schroder's and Harvey's measures of integrative complexity.

The best-known study of generality is Vannoy's (1965) article based on his dissertation (1964). Vannoy includes Schroder's Paragraph Completion Test (PCT), but not Harvey's "This I Believe" (TIB) Test (both the PCT and the TIB are intended to measure integrative complexity), plus a number of other tests of varying and sometimes uncertain relation to complexity. He concluded that complexity is not unitary but comprises two or three uncorrelated components. Unfortunately, the study contains several serious flaws. No one claims that complexity is unitary. As I have noted, most theorists hypothesize components of complexity, and most suggest that complexity will vary with content. Any adequate examination of the generality issue would have to be organized around these points, and Vannoy's failure to keep them in mind has resulted in an interpretive jungle. First, not all of the tests he used are primarily intended to measure complexity. A case can be made that the scores might be related to complexity, but how

much and in what ways is not clear. For some measures, viz., the "social distance" questionnaire, the scores as used (variance components) are absolutely ambiguous. Second, the cognitive domain varies among tests. To the extent that there is a consistent domain, it would be interpersonal, but even this rubric incorporates a broad range of types of stimuli, varying from four adjective descriptions of hypothetical people to real persons nominated by the Ss to fill particular roles. Clearly, some of the measures included should not have been, and some important procedures were left out.

Of most immediate interest here is the finding that Schroder's PCT loads highly on the last of Vannoy's factors, is not accompanied on that factor by any other complexity measure (SCAT-Q is second highest) and shows no other high loading itself. It would seem natural to notice, although Vannoy does not, apparently because he did not attend to such distinctions, that the PCT is the only test in the battery specifically designed for measuring integrative complexity.

A much better study is that of Gardiner (1968). Gardiner's greater sensitivity to "integrative complexity," as espoused by Schroder and Harvey, leads him to include a total of six measures of this variable. Scott's (1962) sorting task is included, but is applied to nations (as Scott used it), so its failure to relate to other tests may be due to independence of complexity between domains as much as to independence between dimensionality (if sorting measures dimensionality) and integrative complexity. No true measure of dimensionality is included (e.g. Carr, 1965). The REP

appears only in a check list "version." Also included in the battery are a number of variables which are expected on theoretical grounds to relate to complexity. In most cases, these derivations are quite straightforward: generally, the theorists themselves have stated that such-and-such a characteristic is related to complexity, and Gardiner has chosen a commonly accepted (or acceptable) measure of the characteristic.

The major findings are, first, that the measures of integrative complexity (the TIB, PCT and Interpersonal Topical Inventory, or ITI, a forced-choice objective version of the PCT, two scores being derived from each of these tests) define the first factor in a factor analysis; and, second, that the third factor, which Gardiner pays less attention to than the second, appears to approximate the idea of a general complexity factor. Tests loading on this factor are the PCT, TIB, Scott's sorting task, and Embedded Figures Test, plus three other tests requiring flexible, analytic thinking. In fact, the label "Analytic Thinking" could be applied to the factor, an interpretation which would seem consistent with Schroder's (1969) description of complex thinking. Gardiner suggests that the factor relates to dimensionality or cognitive differentiation, noting that the PCT and TIB relate to dimensionality, while Scott's nation sorting task is "specifically dimensional." In that case, the loadings belie a strictly dimensional interpretation, since both scoring versions of the TIB, as well as the Embedded Figures, load more highly than does the Scott task.

The third major finding is that, according to Gardiner's original

factor analysis, a number of characteristics which have been hypothesized to covary with complexity do not, including two measures of flexibility, Associations IV and Gestalt Transformations; two tests involving acquiescence, Social Acquiescence and Seeing Deficiencies; a scale of Internal-External Causation; and two measures based on tolerance for inconsistency in attributed descriptions (references for these tests may be found in Gardiner, 1968). However, there is some question as to the appropriateness of testing such hypotheses with factor analysis. The virtue of factor analysis is that it compresses or organizes large quantities of data. In the process, finer details may be lost. (Additionally, of course, there is the basic arbitrariness of the location of the dimensions and thus of the factor loadings.) Thus, whenever possible, hypotheses should be tested without first subjecting data to the leveling effect of factor analysis.

This more direct type of testing is used in examining three hypotheses in the study: (A) that there will be no sex differences in complexity, (B) that firstborn Ss will be less complex than laterborn, and (c) that indiscriminantly pro-religious Ss will be less complex than others. Each hypothesis as stated is based on theories about complexity. There should be no sex differences because no one has ever given grounds for any, or suggested that there would be. Anticipated differences between first and laterborn are based on widespread findings (reported in Gardiner, 1968) that firstborn are more dependent, acquiescent and conforming and, more importantly, that these differences are based on childrearing practices

which differ in such a way, according to the theorizing of Harvey et al. (1961), as to justify the current prediction. The relationship between religious conviction and complexity is based largely on Harvey's (1966) statements and scoring criteria for the TIB. Each hypothesis was tested by dividing Ss into the two appropriate groups and using t tests to compare scores on the other variables.

In each case, the hypothesis was not supported. Females were more complex than males. Gardiner explains this rather plausibly by invoking some class-related differences in parental treatment of boys and girls. The failure to find birth order effects is not well handled, however. Since he argues that the sex differences are due to differences in child-rearing practices, and that the anticipated birth order differences would have been due to child-rearing practices, it may be that the two are connected. For example, if a disproportionate number of firstborn were also females, the hypothesized birth order effects could have been washed out by the hypothesized sex effects.

Finally, indiscriminantly pro-religious (IPR) Ss were found to be more complex than other Ss, a result opposite that expected. Gardiner attempts to explain this first by stating that the IPR Ss were more intrinsically oriented than the non-IPR Ss (a "complex" characteristic), and, second, by pointing out that the Ss were young (15-16 years old) and thus not yet fully differentiated regarding religion; i.e. that the IPR group included both the genuine IPR plus intrinsically religious Ss who are currently

also global but will not be in a few years. While this is not unreasonable, it also avoids some issues. Ss are all of an age, thus differentiation is relative. One would still expect relatively complex Ss to be non-IPR. At worst they might be unrelated, if religion were a distinct domain from that in which complexity is measured. In the present case, however, IPR Ss are more complex. This is particularly unexpected in light of the fact that the measures of complexity which reach significance are based on the TIB scoring criteria which state that IPR behavior represents Level I functioning. It would seem that the hypothesized relationship between religious belief and integrative complexity is not present.

On this basis, then, it would seem that the PCT and the TIB are related, even though Gardiner does not report the correlation between the two. Unfortunately, there is one further piece of evidence which contradicts this finding. Schroder et al. (1967, p. 115) comment virtually en passant that the two tests bear only a "low-order relationship." So the picture is confused.

Summary

Both Schroder and Harvey share the position that integrative complexity is the most important component of cognitive complexity. This makes their position distinguishable from that of the more dimensionally

oriented theorists (Bieri, 1961; Scott, 1969). It also suggests that the most immediate question is whether Schroder and Harvey agree on their measures of integrative complexity, rather than whether integration is related to dimensionality.

Regarding the comparability of the Schroder and Harvey positions and measures, first, there is evidence that in factor analyses the two tests constitute a separate, common factor: Vannoy (1964) found the PCT defining its own factor, and Gardiner (1968) showed that the PCT and TIB (plus the ITT, a derivative of the PCT) all load highly on a factor by themselves. The correlation between the measures is not reported in the Gardiner study, however, and Schroder et al. (1967, p. 115) have stated that the two tests bear "only a low-order relationship." Thus, the evidence is conflicting.

Second, although they are not predicted, sex differences in complexity have been found in at least one case (Gardiner, 1968). These differences were attributed to differences in child-rearing practices for the sexes, but it is not impossible that other, albeit more subtle, sex-related differences in child-rearing practices and/or cultural role expectations would affect complexity at other socioeconomic levels. Thus, the question of whether there are sex differences in complexity is not settled.

There is another area of ambiguity which will be briefly mentioned here. Schroder et al. (1967, p. 197) also state that the correlations between PCT scores and intelligence have ranged between .12 and .45 for various samples. This may remind us that the second highest loading on

Vannoy's "PCT factor" was SCAT-Q, and perhaps also of the analytic cast to Gardiner's (1968) third "general complexity" factor. (One of Gardiner's greatest deficiencies was his failure to include an intelligence measure.) The general question is, then, what is the relationship between intelligence and integrative complexity?

A somewhat related question is that of sample characteristics. Not only are Ss typically college students, but college students from fairly unrepresentative schools. The potential effects of these biases have apparently never been fully considered.

Before further developing these points, some background discussion of theory seems appropriate.

Chapter II

CONCEPTUAL SYSTEMS: THEORY

The fullest description of this theoretical position is found in Harvey, Hunt and Schroder's (1961) book. The basic ideas from this source appear in briefer, more readable form in Harvey and Schroder (1963), Harvey (1966), and to some degree in Schroder and Harvey (1963). In general, it appears that Harvey continues to espouse an essentially unchanged version of the 1961 theory, whereas Schroder harbors some revisionist tendencies, hints of which may be noted in the Schroder and Harvey (1963) paper.

The Theory of Conceptual Systems: Harvey, Hunt and Schroder

Conceptual systems are composed of concepts and their interrelations. The definition of "concept" varies somewhat from place to place, at one point being defined so as to imply a dimensional characteristic (". . . a system of ordering that serves as the mediating link between the input side

[stimulus] and the output side [response]" p. 1[1]), in another case suggesting that concepts are points along dimensions (concepts represent "a category of varying definitiveness and breadth along some specifiable dimension. . . ." p. 10), and in still another situation saying that a concept is minimally a comparison of two points on a dimension, and may be a point or points in a space of infinite dimensionality ("The simplest concept is one involving two points on a single dimension, and the most complex concept is one embodying an infinity of points on an infinity of dimensions." p. 62). These differences are probably not critical, although the last seems different as well as overdrawn. The basic idea is apparently quite similar to George Kelly's (Kelly, 1955); namely, that organization is imposed on perceptual experience, that "meaning" (or "understanding") derives from the imposed organization, and that this organization is in varying degrees idiosyncratic and arbitrary. Concepts ("constructs" and "contrasts," in Kelly's terminology) are the terms in which the world is viewed; that is, concepts are the content of conceptual systems. Harvey, Hunt and Schroder wish to regard differentiation-integration as a process, and concreteness-abstraction as a result of the process. Consequently, neither of these dimensions is inherent in the stimuli, but rather in the conceptualization of the stimuli.

An increasing number of concepts implies increasing differentiation

---

[1]In this section, page numbers refer to Harvey, Hunt and Schroder, 1961.

on the part of the person, but greater abstractness is not an automatic concomitant of greater differentiation. True abstractness requires integration of concepts as well as differentiation, and differentiation may occur without subsequent integration. Also, although differentiation is, in some sense, a logical prerequisite to integration, these authors argue that the two can occur essentially simultaneously. Unfortunately, important as it clearly is to the theory, the nature of integration is never really made clear. "Integration is the relating or hooking of . . . parts to each other and to previous conceptual standards" (p. 18). A straightforward interpretation would be that differentiation involves seeing differences and integration involves seeing similarities, both of which can be along dimensions. However, the book consistently suggests, and there are explicit statements elsewhere (e.g. Schroder et al., 1967, p. 7), that dimensions and integrative rules are qualitatively different. This failure to clarify the integration concept is a near fatal weakness in the theory, and one which persists in the current positions.

Given proper parental training practices (to be discussed below), children will progress from concrete to abstract functioning. If conditions are unfavorable, arrestation will occur at some point along the dimension, the stopping point being determined by the nature of the unfavorableness. Of course, the level to which a person progresses is likely to vary from one content domain to another, again as a function of varying training conditions. Harvey, Hunt and Schroder have concentrated on the interpersonal domain,

which is appropriate, but how and the degree to which their discussion would generalize to a more content-oriented domain (e.g. "history") is not always clear.

In the theory, introduction into a new, unfamiliar realm of experience necessitates relatively concrete behavior, i.e. dependency on external sources or authority. With increasing familiarity the potential for increasingly abstract behavior develops. The authors cite a number of developmental theorists in support of their argument that children typically pass through several concrete to abstract cycles, the concrete phases corresponding to the times at which children begin to talk (roughly age 2), begin social behavior (age 5), enter peer-centered society (age 10), and begin to take on adult roles (apparently seen as about age 17). The first cycle progresses in the direction of abstraction only as far as the second stage or level, or perhaps to the transition between the second and third levels. Later cycles progress as far as Level IV.

The first stage (Level I) is labeled "unilateral dependence." It might also be termed a "pro-authoritarian stage": the connotative baggage seems to be the same (provided emphasis is on style or process, not content). If differentiation and integration occur to a proper degree, Level II will be reached,[1] which is described as a period of negative independence, and in

---

[1]Much of the elaborateness of the book, which is considerable, is due to extensive consideration not only of the major stages but also of the transitions between stages. This material seems to have dropped out later and will not be considered here.

at least some forms bears a strong resemblance to anti-authoritarianism. The categoricalness is similar to Stage I. The distinction between the levels is on the basis that at Level II the person is differentiating himself from the external authority, whereas at Level I he is defining himself in terms of it. Exploring this differentiation takes the form of resistance to the authority. Integration of the two possibilities, necessary for progress, would make available to the person some form of both attitudes and both types of behavior. Fixation at this level, a failure to integrate, would leave only a strong negativism.

Level III is characterized by conditional dependence and mutuality. In this stage, hypothetical or conditional ("as-if") thinking becomes possible, and a more objective, information-oriented attitude is taken toward the world. Socially the person at this level of abstraction is able to see another's point of view and to understand the other person from that point of view. Additionally, the third stage permits the "holding of alternative views of the self, of events, and of others simultaneously with a minimum of concern for ambiguity" (p. 101). By implication there is still an element of dependency and/or a diffuseness in the self-concept, because Level IV is defined as the stage at which ". . . mutuality and autonomy are integrated so that neither interferes with the other and yet both are important" (p. 106). In science and other noninterpersonal endeavors, Level III appears as an empirical orientation; so presumably IV is the theoretical stage, although the analogy with mutuality and autonomy does not appear to be strong.

Level IV functioning is characterized by an increased sense of internal causality as well as increasing problem-solving ability, tolerance of anxiety, ambiguity and inconsistency, greater resistance to stress, and a stronger orientation toward informational solutions to questions or problems.

If training conditions are unfavorable, then development will be arrested at some point along the scale, the point of arrestation being determined by the nature of the training practices. Again, a super-articulated theory will be greatly simplified, for two reasons: to conserve space, and because the theorists themselves have apparently paid no serious attention to the elaborations in their own research.

Training is of two major types, unilateral and interdependent. Each of these, in turn, has two subtypes. Thus, the four major types of training are (1) reliable unilateral, (2) unreliable unilateral, (3) protective interdependent, or (4) informational interdependent. These dispose toward arrestation of the trainee at Levels I through IV, respectively.

Reliable unilateral training implies an authoritarian orientation in the trainer. That is, the criteria for rewards and punishments are external, "objective," rigid, evaluative and absolute. Good behavior, as defined by the criteria, is consistently rewarded, and bad is consistently punished. There are few, if any, lapses in either direction. Questioning of the criteria is not permitted. Requirements, however, must, in some sense, be within the capacity of the trainee. Training of this sort is hypo-

thesized to lead to Level I functioning.

Unreliable unilateral training (defined from the trainee's point of view) occurs primarily either when the trainee is unable to discover the external criteria (e.g. when feedback is too erratic) or when he is unable to conform to it (as when the demands are beyond his ability). In these instances, the trainee either sees no pattern of reward and punishment, or builds up an expectation of failure. A third condition which may be experienced as unreliable occurs if parents are indifferent toward and neglectful of their children. Fixation at Level II is likely to follow from any of these circumstances.

Protective interdependent training is based on systematic rewards for instrumental behavior, where the goal of the behavior is not evaluated against external criteria as in the reliable unilateral condition. In this case, support is used both as a reward and as a means of directing behavior. This support is likely to be so extreme as to become a tendency to protect the trainee, insofar as the agent is capable, from failure experiences. Thus, although relativism and an emphasis on interaction exist, there is also a restriction as to the kinds of experiences the trainee is permitted, and progress beyond Level III will probably not occur.

Under informational interdependent training, the trainee is permitted to learn from the consequences of his own experiences, including failure, with the important proviso that the environment be so structured that the subject is not pushed beyond his capacities. So long as this is the case,

the trainer feels, protection is unnecessary and failure is as informative as success. It is this informational aspect of experience which is emphasized and valued by the trainer. Such an approach permits the development of the abstract skills required for Level IV functioning.

For present purposes, this is an adequate summary of the 1961 position. Modifications in the theory will be considered below; however, it should be emphasized that what Harvey, Hunt and Schroder set out to do was to describe a cognitive variable which is independent of content but which would predict certain styles of behavior. Seemingly, they have been successful at a descriptive level at least. For example, theoretically the authoritarianism of Level I may be of the left or right, pro- or anti-religious, behavioristic, humanistic, etc., and supposedly the same is true for each of the other levels. As we will see, Harvey has moved toward a closer relationship with content, and Schroder has moved further away.

Before discussing such changes, however, a problem inherent in the development and training arguments should be mentioned. Basic to the problem is the fact that by either measurement technique (to be discussed below) complex Ss will comprise only 4% to 10% of a sample ("complex" has different meanings--Level IV for Harvey and approximately a quarter of the way between II and III for Schroder). That is, in young adult (college) populations, complex people are quite rare. This would seem consistent with the description of the training agency hypothetically required to pro-

duce abstract thinking. It is not consistent with the discussion of developmental cycles, in which it is stated that the modal (i.e. most common) behavior for nine-year-olds is at the level of transition into the fourth stage, and that Level IV functioning "is more likely in the adolescent cycle of development." That is, at age 9 or 10 the most common functioning is very nearly the most abstract possible, but by age 20 this same abstract behavior has become the least common found. My suspicion is that the theorists were carried away in their desire to integrate a variety of source materials, although there is at least one alternative explanation; namely, that parents are relatively interdependent and schools relatively unilateral trainers, and that by college the increased proportion of time spent in school, plus selection, have done their job. This hypothesis seems unlikely, however.

Conceptual Systems: Theoretical Modifications

No significant theoretical revisions are apparent in Harvey's more recent writings. The most recent exposition of which I am aware (for example, Harvey, 1966) is a concise, convenient summary heavily based on Harvey et al. (1961). (The tendency toward closer ties to content will be discussed in conjunction with measurement.) There is reduced confusion over the term "integration," but this results not from elaboration of a con-

cept which is different from dimensionality, but from a simplification of the concept until it becomes itself essentially dimensional. Integration is said to be "interrelating," which can easily be given a dimensional interpretation. This, in turn, means that there has become little distinction between Harvey's theoretical position and a strictly dimensional definition of complexity (e.g. Bieri, 1966).

Schroder (1969) stands in some contrast to Harvey, since a number of changes have been introduced into his theorizing, most of them having to do with the nature of the concrete-abstract dimension.

First, specification of the dimension has even less to do with content or with behavioral styles than in earlier material. The dimension is conceived in more truly structural terms. Schroder no longer specifies the nature of the poles of the developing dimensions, and the specification of the behavioral characteristics is more obviously derived from the structure of the cognition as opposed to the nature of the dimensional poles. Further, the model is far more easily generalizable to domains other than interpersonal, and the progression through the stages is more logical and orderly, and more easily seen as developmental. All of this has strengthened the model enormously.

Second, the ideas of differentiation and integration have been amplified, and in the process definitions have changed. Finally, there is a strong tendency toward specification of three instead of four levels. This is probably for practical reasons, as virtually no one operates even at

Level III on the device used by Schroder to measure conceptual level.

More concretely (if that can be permitted) and briefly, Schroder's current position (Schroder, 1969) is as follows: Differentiation is now defined as "the number of categories or kinds of information extracted from a given domain of stimuli," or the number of unique principles of stimulus ordering, or the number of scales or dimensions used by the person in viewing stimuli. These scale or dimensional values are the informational input, and for a given stimulus the amount of information will vary from person to person with the degree of differentiation. It is hypothesized that people learn to combine these values in various ways according to learned rules. The integrative complexity of the person's conceptual structure is determined by the number of these combinatory rules and by the "degree of connectedness" between the rules. A combinatory rule is a pattern of weights assigned to the scale values, so if stimuli are scaled along three dimensions, and the values along the scales are weighed 50, 30, and 20, then the resultant values have been arrived at by a combinatory rule (alternately termed a "perspective"). Level I functioning occurs when only one such combinatory rule, or set of weights, is available for a given set of scales. In simplest terms, this means that a concrete person is one who can see only one alternative or one conclusion in a particular situation. When two or more alternatives, patterns of weights, understandings, perspectives, interpretations, judgments, or whatever, are available, simultaneously and for the same information, then the person is operating at

Level II. Connectedness between these two or more perspectives plays the same role as integration or integrative rules in Harvey, Hunt and Schroder (1961), and is the necessary condition for Level III functioning, the highest Schroder considers. "Connectedness" is every bit as poorly defined as "integrative rules," which is the most obvious and immediate criticism of the theory. To connect perspectives is to relate perspectives, and beyond this the concept is simply not developed. My best guess is that connecting perspectives means to compare them against some superordinate, relative construct, like goals, values, or motives.

In spite of the terrific flaw resulting from Schroder's failure or inability to specify the nature of connectedness more fully, I find his position intuitively and aesthetically more attractive than Harvey's. This is not to say that there are not other problems, or that there are not viable alternatives.

It seems clearly to be the case that perspectives, the results of combinatorial rules, may be regarded as dimensions. Each perspective orders stimuli in a new way, which is the definition of a dimension. All that is necessary to locate the new dimension/perspectives is to find an axis such that the cosine of its angle with each of the dimensional axes is equal to the weight given ratings on that dimension, expressed as a decimal fraction. There may be cases where this would not be possible, but, in general, I suspect it could be managed.

I can imagine two objections Schroder might raise. First, the new

factor would be oblique, and he specifies that informational dimensions be mutually orthogonal. The urgency of this requirement is not clear to me, inasmuch as orthogonal dimensions are probably rare at best, and may be psychometric inventions. Oblique relationships are probably, at the very least, more common. (Schroder uses "friendliness" as an example of a dimension: to what is the friendliness dimension orthogonal?) It is not even clear that research reported later in the chapter conforms to this requirement. In one study, stimulus persons were described along three dimensions: creativity, orderliness, and reciprocity, which Schroder reports as being independent. The sense in which they are independent is not clear (statistically or cognitively), and it may well be that at least some Ss would feel moderate dissonance if a target were described as very creative and also very orderly.

The second objection to viewing perspectives as dimensions might be based on the statement that differentiation is the "kinds of information extracted from a given domain of stimuli." The implication is that information is somehow intrinsic in the stimuli, whereas perspectives are the result of cognitive processing. I regard this as an untenable distinction. Examples of dimensions used by Schroder have been mentioned above (friendliness, creativity, orderliness, reciprocity). The only case in which these are "given" occurs when an experimenter tells a subject that a stimulus person is "friendly," etc., and even in this situation the connotations of this information will vary among Ss. Normally, making judgments along

dimensions will involve processes like those involved in perspectives. The differences, if any, will have to do with the ease and/or speed with which the process operates. In general, perspectives and dimensions seem comparable.

Because the nature of connectedness is so unclear, it is more difficult to attempt a dimensional reinterpretation. However, if "connectedness" means "relating," this suggests to me the finding of similarities and differences, which, in turn, suggests dimensions or concepts (equivalent in Schroder's terminology). If connectedness is in terms of goals, values or motives (one's own or others, simultaneously or one at a time), I suspect these could be represented as idealized dimensions or points in the space, against which comparisons could be made.

This predilection toward a dimensional model demonstrated above has two bases; first, a reaction against the Schroder and Harvey claim that they are dealing in concepts which dimensional approaches cannot explain, especially since the integration concept is never specified, and, second, an interest on my part in pushing the dimensional model to its limits before introducing additional constructs. However, even if this dimensional reinterpretation of conceptual levels is plausible, Schroder is raising an important point, and one which a dimensional approach may have trouble representing; namely, the issue of super- and subordinate relationships among dimensions or perspectives. Kelly (1955) identified two types of such hierarchical arrangements, one in which the dimensions lie, as it

were, parallel to each other (e.g. "intelligent-stupid" is in some degree subordinate to "good-bad"), the other in which both poles of one dimension are subordinate to one pole of the superordinate dimension (Kelly's example: "light-dark" are both descriptive on a "descriptive-evaluative" dimension). These examples could probably be handled by leaving empty quadrants in the space, but this seems inelegant and the general issue could quite possibly prove thorny.

Schroder's connective rules, then, would probably have to be viewed as superordinate dimensions if they are to be dimensions at all. Abstract functioning, in Schroder's sense, would be dependent upon the development and use of such superordinate dimensions. Conceptual systems theorists have argued that abstraction is not related to dimensionality. This would be an agreeable statement so long as the dimensions considered are at a given ordinal level. In passing it should be noted that the REP test (Kelly, 1955), commonly used to measure a dimensional form of cognitive complexity (e.g. Bieri, 1966), is not used to distinguish super- and subordination. Some dimensional measure is needed which will reveal such structure before it can be determined whether the relationship proposed above exists.

A comparison of the current theoretical positions of Schroder and Harvey will be delayed until their respective measures have been considered.

Chapter III

CONCEPTUAL SYSTEMS: OPERATIONS

Although a number of procedures are viewed as yielding scores relevant to structural complexity, in practice both Schroder and Harvey have evolved particular tests which they regard as criterion measures. The two tests are both similar and different and will be discussed in this chapter.

Schroder's Paragraph Completion Test (Schroder, Driver & Streufert, 1967)

The Paragraph Completion Test (PCT, Appendix I) presents the subject with a word or phrase which is supposed to be the beginning of a sentence, and requests S to finish the sentence and write two or three further sentences on the same topic. There are five such items, plus a warm-up, nonscored item. Each item is timed, the limit ranging from 100 to 130 seconds.

Items are rated by trained judges, scores being given on a 7-point

scale, the odd numbers representing the four major levels. The mean (or sum) of the two highest item scores is the preferred overall score. Inter-rater reliabilities for this score reportedly range from .80 to .95, with inter-item correlations mostly in the .40s, ranging between .39 and .62. No test-retest reliabilities are given.

Scoring criteria are as follows (based on Schroder et al., 1967, Appendix II): A score of 1 ("Low Integration Index" or Level I) is given for a response generated by a single rule, where alternatives are neither present nor considered, and where thinking is categorical and without conditions. Additionally, (a) conflict is viewed as bad, and uncertainty or ambiguity as indicating weakness; (b) S seeks fast, unambiguous closure via internally consistent, balanced processes; (c) a specific rule or guide is offered which leads to or implies an absolute, categorical solution; (d) S compartmentalizes or categorizes facts and effects and tends to treat only one side of an issue.

Score 2: An absolute score is qualified, but no clear alternative is given.

Score 3: (Medium-Low Integration Index, Level II): Two or more "alternative rule structures" (or perspectives or interpretations) for viewing the event are simultaneously present, plus possibly a conditional rule for selecting among the alternatives. Indicators include (a) mentioning two or more different interpretations of events; (b) listing similarities or differences without considering relationships; (c) giving conditional rules for

choosing among ways of categorizing; (d) reacting against absolutism and/or avoiding dependence on external forces.

Score 4: Alternatives are offered, with the implication of interaction, but the "interaction is expressed as a qualification rather than as the emergence of comparison rules."

Score 5: (Medium-High Integration Index, Level III): Alternative interpretations are given and interact in some specified way. More specifically, (a) the alternatives are related in a way which maintains the ambiguity or conflict; (b) the existence of the alternatives is itself regarded as having various meanings; (c) other people's points of view are taken into consideration; (d) indication that one's behavior is influenced by others' behavior and/or that social interactions should be based on mutual responsibility; (e) seeing alternative reasons for similarities and differences in perspectives.

Criteria for scores of 6 and 7 (7 is the High Integration Index, Level IV) are described, but such responses are exceedingly rare and will not be considered here.

It should not be difficult to achieve the interrater reliabilities Schroder reports. The criteria for scores above 5 are extremely abstract, but since such responses are rarely encountered there is no great problem. One difficulty would be in discriminating between "polarity responses" (Score 1) and true alternatives (Score 3). The former are alternative decisions based on a single rule (e.g. "good or bad"), whereas the latter are

based on different rules (something might be good and bad simultaneously). Once these can be separated, rater reliability should not be hard to establish.

There is some potential confusion about the distribution of scores, however. Schroder et al. (1967, p. 195) indicate that, for the most concrete and most abstract of their samples to date, approximately 4% and 12% of the Ss (respectively) obtained scores of 7 on a 7-point scale, i.e. obtained the highest possible score. At that time a S's score was either the mean of all his item scores or the mean of the two highest item scores. Current practice is to add the two highest item scores together to decide a score, so these are distributed between 2 and 14 (a 13-point scale). The confusion lies in the fact that the scores on this 13-point scale are seldom higher than 7 and that this distribution of scores on the lower half of the 13-point scale (the lower half of the complexity dimension) may be mistaken for the distribution reported in Schroder et al. for the entire 7-point scale (the whole dimension).

Schroder explains this shift in distributions[1] by stating that there has been a change in scoring procedures (the changes appear to be in the application of the scoring criteria, not in the criteria per se), so that now raters are being especially careful to avoid making liberal assumptions about or interpretations of responses, since these were leading to unreliable dis-

---

[1] H. M. Schroder, personal communication.

tinctions in the upper range of scores.

The implications of the shift are unclear. The "new" distribution is more nearly normal than the "old" one, i.e. it is sharply peaked in the center and has few scores at either extreme. The old distribution was much flatter, and in some samples, the mode was the lowest, most concrete score. The new scores are obviously not merely a scaling-down of the old distribution. Thus, the relationship between complex Ss as defined by Schroder et al. (1967) and as they are used today is not certain. It is not clear what score was required to constitute complexity in Schroder et al. but, by present standards, a complex S is one for whom the sum of his two highest item scores is at least 7 (3.5 on the 7-point scale); that is, very slightly above Schroder's "Medium-Low Integration Index," or very slightly above Level II in Harvey (1966) or Harvey, Hunt and Schroder (1961) terminology (ignoring theoretical differences). Put differently, a complex person, by Schroder's current definition, is one who sees alternative interpretations to events, and who additionally shows the tiniest wisp of ability to integrate interpretations. It seems likely that, if anything, complexity so defined is even less common than complexity as defined by Schroder et al. (1967).

Obviously there is nothing wrong with being interested in uncommon characteristics, but such an interest does introduce certain problems. In the present case, the theory becomes virtually inapplicable to naturally occurring groups (a problem exacerbated by the restricted variance of the

scale and by Schroder's habit of using extreme samples and ignoring the middle group). Further, the reliability of the test becomes extremely important.[1] If the test misidentifies subjects, considerable mischief can be done. Given this fact, it is especially odd that Schroder does not report test-retest reliabilities.

Insofar as validity is concerned, the test does appear to make an important distinction among Ss regarding their information-processing and decision-making capacities. Schroder's typical procedure is to test a large enough pool of Ss to obtain a given number of "complex" persons. "Simple" Ss are then chosen who match the complex Ss on important control variables (intelligence, in particular, but sometimes other variables also). Groups of Ss so defined differ in the amount of information they are able to utilize in decision-making, and in the quality of the decisions made, the complex groups being more coordinated, interrelated and self-directed. The tasks to which the groups apply themselves vary, but are characterized as being long term (usually several hours overall), relatively realistic and interactive such that what a group does affects what happens later.

There is an interesting anomaly here. Schroder feels that his PCT measures interpersonal complexity, and he exhibits the usual caution about generalizing to other domains. Yet the games or tasks which the groups are given are not interpersonal in nature, thus there would not appear to be

---

[1] I am indebted to Lawrence Stricker for pointing this problem out to me.

any basis for predicting differences between the performances of interpersonally simple and complex groups. Schroder obviates this objection by arguing that the content of the game is not so important as the fact that decisions must be made in a group, so interpersonal dynamics are important, and the PCT again becomes relevant. One study has been run using single subjects. In this case, complex Ss were still superior to simple, but here the task had to do with learning about a culture and solving a problem based on one's knowledge about and understanding of that culture. In this case, Schroder argues that the task is analogous to interpersonal situations, in that similar skills are required for success in both situations. So far as I know, individuals have never been run on the noninterpersonal games.

In summary, the PCT is a sentence completion measure of integrative complexity. It has satisfactory interrater reliabilities, and no obvious serious problems in the rating process. Test-retest reliability is not established; however, extreme groups, matched for intelligence (a correlate of PCT scores) do differ systematically in predicted ways on information-processing and decision-making tasks. One problem with the test is that complex persons are rare, which emphasizes the importance of the missing test-retest correlation, and which limits the applicability of the measure.

Harvey's "This I Believe" Test (Harvey, 1966)

The "This I Believe" Test (TIB) is similar in format to the PCT. Ss are presented with a series of sentences reading, "This I believe about . . .," the blank being filled in by such referents as the American way of life, marriage, religion, friendship, etc. (see the Appendix). Items are timed, and Ss are requested to write at least three sentences. Generally, about 10 items are used.

Trained raters assign scores to the protocols. However, individual items are not scored. Rather, the entire test is given one, overall global, score. An attempt is made to place each S at one or another of the four levels. Some protocols are mixed, however, and functioning at more than one level is indicated. In these cases, two (or even three) scores may be assigned.

Harvey states (1966, p. 47) that the "inter-judge reliability for three and four trained judges in classifying subjects into one of the four main systems has been .90 or above for 12 different samples of subjects." I take this to mean that, for the Ss for whom each judge gives a single assignment, interrater reliabilities equal or exceed .90, and that Ss given more than one rating do not appear in these correlations. Harvey reports that, out of 1400 TIBT's, 30% were System 1, 15% System 2, 20% System 3, 7% System 4, and the rest (28%) were mixed.

No test-retest reliability is reported, so far as I have been able to

discover.

Scoring criteria can best be described as messy. They are complex, rather content-oriented (theory to the contrary notwithstanding), and multi-dimensional in nature. For example, Harvey describes some of the more important criteria as,

> . . . the absolutism of his expressed beliefs, consideration of contingencies or modifying circumstances, dependency on external authorities, especially God and/or religion, frequency of trite and normative statements, degree of ethnocentrism, acceptance of socially approved modes of behavior, concern with interpersonal relationships, and the apparent simplicity-complexity of the interpretations of the world (Harvey, 1966, p. 47).

More specifically (Harvey, undated), System 1 Ss tend to be absolute and closed in their thoughts and beliefs, to be highly evaluative and conventional, to show a strong positive attitude toward traditional authorities, to be highly ethnocentric, and to identify closely with social roles, status requirements, etc. These Ss appear to have internal standards, but these derive from reliance on external authority. This orientation may be characterized as pro-authoritarian.

System 2 Ss are anti-authoritarian. They dislike and distrust authority and automatically rebel against norms and social pressures. They are as rigid, simple and consistent as System 1 Ss, but in the opposite direction. They add hostility, rejection and negativity to the concrete, non-analytic, categorical thinking of System 1. Harvey also states that, while these Ss reject conventional standards, they are simultaneously without other stable guidelines. They lack personally derived standards, and

Harvey does not mention (but implicitly rejects) the possibility of subgroup norms.

System 3 Ss appear to be high in need affiliation. They are socially oriented and friendly. They tend to be dependent on others rather than on their own resources in the solution of their problems. They are quite acquiescent in a socially accommodative sort of way. Their opinions tend to be flexible and superficial. They feel that everyone needs lots of friends. The hypothesized increase in complexity over Systems 1 and 2 is based on the assumption that these Ss have learned more, and more sophisticated, techniques for dealing with other people (System 3's, after all, are manipulative--necessary if they are to be dependent), and are therefore more goal-oriented and more flexible.

Unless it is "inner-directed," no familiar label exists for System 4 Ss. They are supposed to be independent, flexible, open, nonevaluative, nonjudgmental, and relative. They like novelty, and are relatively able to tolerate ambiguity and/or dissonance. They are "information-oriented," which is to say, rational, analytical, and empirical. They tend to present and juxtapose diverse elements in their responses, and to evaluate these noncategorically and relativistically.

These criteria are, categorically, difficult to apply, for the reason that the criteria within a system are not necessarily highly correlated. This means that it is not uncommon to get some indicators but not others, to get an indicator that does not distinguish between two systems, or worse,

to get indicators from more than one system. For example, if a respondent sees no value in (condemns) the American Way of Life (Level 2) but simultaneously argues that people should concentrate on learning how to communicate better with each other (not System 2, because an alternative value system is proposed), should the score be 2 or 4?

A number of other rules exists for scoring, having to do, for example, with the overall tone (pessimistic-optimistic) of the response, the attitude toward friends, etc. The effect of these rules is to lead to difficulty, for this rater at least, in discriminating between 1's and 3's, on the one hand, and 2's and 4's on the other. If the decision is at all close, the best strategy, if one is attempting to establish rater reliability, is to give both scores and thus remove the S from the pool. On such ambiguous protocols it is not particularly difficult to obtain moderate negative correlations between raters. If Harvey is talking about a dimension, confusion should occur between adjacent categories, not alternate ones. If assignations jump about as described above, and in my experience there is this tendency, then the unidimensionality of Harvey's four systems seems questionable.

Also, it should be noted that there is considerable dependence on content in the scoring. A strong positive attitude toward religion and/or other institutions constitutes System 1, negative attitudes System 2. Someone who feels that a lot of friends are important is a System 3, but if friendship is unreliable he is System 2. Examples could be easily added. In practice, only with System 4 criteria do structural considerations (how something is

said rather than what is said) begin to outweigh content. I would suggest that for the first three systems the TIB primarily measures content rather than structure, and, in fact, measures structure only insofar as it is related to content.

Of course, Harvey does see a relationship between content and structure, and the relationship is based on theory. In the process he reveals his theoretical conservatism, because the derivation of the relationship can be found (perhaps by implication for some specifics) in Harvey, Hunt and Schroder (1961). The original theory specifies a universal sequence (at least in this culture) in which interpersonal experience is differentiated and integrated, including in a general way the nature of the developing dimension at each stage. This, in turn, implies something about the content of attitudes and behavior, in the sense of the focus of interest and the general style of approach to the topic. It is on this specification that Harvey draws for his scoring criteria. The circularity of this process requires that the validating evidence be particularly sound.

Parenthetically, it should be noted that Harvey does not report results as correlations. Like Schroder, he picks groups of Ss who operate at various levels. Unlike Schroder, his Ss are chosen from the full range of his variable. Instead of "simple" and "complex" groups, Harvey usually uses equal numbers of Ss from each of the four systems (levels). These groups are then compared either using *t* tests or analyses of variance. (Of course, the latter would be preferred, but Harvey does not mention

which is used.) Typically, nonlinear results are found; e.g. Systems 2 and 4 are higher or lower than 1 and 3. The effect of this is, first, to support Harvey's rejection of correlation in favor of group comparisons (which should be by analysis of variance); and second, to again raise the question of whether the four systems should be viewed as falling along a single continuum.

Harvey (1966) summarizes a great deal of evidence intended to validate the TIB. Most of these reports fall into one of two groups. The first group constitute trivial findings, theoretically speaking. At best these could be regarded as validating the scoring criteria, and it seems much more appropriate to regard these as hetero-method reliabilities. These relationships should obtain because they are built into the rating criteria. I include here the findings on authoritarianism, dogmatism, need for affiliation, most of the religion material, EPPS, Scott's Scale of Values, and "identification with the American motif." If these had not proved agreeable, then the application of the rating criteria would be under question.

There is a small subgroup which does seem to be of a construct validity nature. First, the WAIS has been administered, and differences between the groups are found only on Verbal Intelligence and Vocabulary, Systems 4 and 2 being significantly higher, although Harvey claims that the systems do not vary in vocabulary usage on the TIB. Of course, it is desirable that the TIB and PCT be independent of intelligence; however, this finding is at variance with Schroder on two counts. First, Schroder re-

ports that intelligence typically correlates with integrative complexity at some level between .2 and .5, a linear, as opposed to Harvey's nonlinear, finding. Second, Schroder has argued (Schroder et al., 1967) that of the components of intelligence the one which should be least related to complexity is vocabulary, because it is based largely on memory and no cognitive processing at all is required for high scores.

A second validating relationship is between cognitive complexity as measured by "a modification" of Kelly's REP test. In this case, a linear relationship was found between complexity and the four systems, System 4 being most complex. Note that this substantiates the position taken earlier that Harvey's is now essentially a dimensional theory, and that it weakens any argument he might make to the effect that his test measures something beyond dimensionality. It also serves the valuable function of ordering the four systems sequentially, and is very nearly the only variable to do so.

The third construct validity relationship is with rigidity. This measure is based on the Gough and Sanford Scale on Rigidity, with which I am not acquainted, and the reference for which is unpublished. Consequently, I can only report that this scale also ordered the systems sequentially, System 1 being the most rigid.

The second major group of validating studies is more impressive. Here anticipated behavioral correlates of complexity were examined. The reports are exceedingly brief and cannot really be adequately evaluated. Tasks intending to measure a variety of learning or concept formation or

cue utilization skills, ability to change set, creativity (novelty plus appropriateness), admission and tolerance of deviant inputs, role playing, etc., were involved. In general, Systems 4 and 1 behave appropriately, with 2 and 3 varying back and forth in the middle. There are some hints that the studies are not independent, the same Ss being used on various tests, and/or the same test-data being analyzed in several different ways. Also, no mention is made of controls for intelligence, which definitely should have been included. If a fuller report of the work met these objections, however, support would accrue for the theoretical position.

To summarize, the TIB relies on scoring criteria which are not fully satisfactory, partly because they are too heavily dependent on content indicators for one to be convinced that the test measures structural properties, and partly because the criteria defining a particular system do not seem to constitute an especially cohesive characterization. The latter leads to problems in assigning Ss to systems, and causes a fairly high disqualification rate (between one-fourth and one-third of the Ss). Validational evidence is spotty. In some cases positive results are built into the scoring criteria. Other predictive validity findings are more impressive, but suffer from the brevity of description and from the implied possibility of lack of independence and control for intelligence. Beyond this, the TIB apparently is related to, but not correlated with, intelligence, and is correlated with Bieri's type of cognitive complexity (Bieri, 1966). On other variables, more often than not, there is a reversal in the sequence such

that Level 2 is more similar to 4, and Level 3 to Level 1, rather than scores being sequentially ordered. Harvey is superior to Schroder in that his scores are more evenly spread along his dimension, and that he considers the full dimension in his research rather than the extremes.

A Comparison of the Two Positions

At a theoretical level, both of Harvey's first two systems seem to represent Level I functioning for Schroder. For example, a System 1 response might be "Law and order must be maintained at all costs," whereas a System 2 S might state, "Cops stink." Each of these represents a single perspective, and so would be scored 1 by Schroder. Harvey's System 4 sounds a good bit like Schroder's score of 3 (Level II), in that a critical characteristic for both is the ability to see alternative points of view (although Harvey seems to emphasize more the development of internal standards). It is interesting to note that the incidence of Harvey's System 4 (7%) is roughly equivalent to that of Schroder's complex Ss (4% to 12%, where "complex" is defined as very slightly above a score of 3, and where a score of 3 represents Level II). Harvey's System 3 has no obvious correspondence with Schroder's scoring system, but it would not be likely to be scored higher than 2, i.e. it would still be simple in Schroder's thinking.

If reasonable, this comparison suggests that the two men are dealing

with different portions of the simple-complex continuum, Harvey's dimension being a portion of Schroder's. Harvey, presumably, is then spreading out his segment, i.e. making finer discriminations within it, than does Schroder. Any such comparison, however, must be in spite of the fact that the scoring criteria for the two measures are radically different.

A second important difference between the theorists seems to have to do with their respective visions of the purpose or value of their variables, or with the way in which they manifest themselves in behavior. I think Harvey wants to identify a characteristic with broad implications, one which affects a good bit of the range of daily behavior, a variable which, when a person's standing on it is known, will convey considerable knowledge as to his typical style. Schroder, on the other hand, is primarily attending to information processing and decision making in complex situations. The kind of complexity with which he is concerned is not engaged in typical behavior, and he goes to considerable effort to devise tasks which are demanding enough that complexity as he views it will begin to play a role.

The last major difference we will consider is that of scoring criteria, where differences are so striking as to scarcely need reiteration. In brief, distinctions among Harvey's first three systems are based on content, whereas Schroder can legitimately claim to be oriented toward structure. This criterion issue reflects on theory, in that it results from Harvey's reliance on a theory which essentially specifies content, while Schroder's restricts itself to structural considerations. This, in turn, means that

Schroder's position is more easily adapted to new circumstances and is aesthetically more attractive in adaption than Harvey's, since the latter's is so intimately associated with the interpersonal domain that one must either refer back to the interpersonal content (in the domain "History," one who relies on the writings of famous historians for his opinions is System 1) or attempt to analogize from it (in "History," what corresponds to need affiliative, that is Level 3, behavior?).

Apparently, then, Schroder and Harvey vary in their definition of complexity and the dimension of complexity; they vary in their research interests and thus in their choice of dependent variables; and they vary in both the definition of scoring criteria for their tests and in the nature of the theories underlying their positions.

The Present Study

The purposes of the present study are as follows:

A. To compare TIB and PCT scores across a broad range of Ss. The implication of the discussion above is that the two sets of scores should be nearly independent.

B. To compare both scores with measures of intelligence. Harvey reports that TIB scores are related to Vocabulary and to Verbal IQ. Schroder states that the PCT should not be correlated with vocabulary, but

should be correlated with components of intelligence involving complex problem solving. A test of the former, and possibly of the latter, is possible with the data at hand.

C. To compare males and females on complexity. Neither theory discusses sex differences, and, so far as I know, neither researcher has looked for such differences. Gardiner (1968) found girls to be more complex than boys, which he explained in terms of child-rearing practices.

D. To compare samples drawn from different populations. Most of Schroder's work is done at Princeton, and much of Harvey's at Colorado. The present study will sample from populations rather different from those typically used by Schroder and Harvey.

Chapter IV

METHOD

Subjects

Approximately 75 female and 75 male subjects were sought from each of three schools. Johnson C. Smith University (Smith) is a private Negro college of about 1200 students drawn largely from the South. SAT scores typically average around 350. Data were collected from 72 women and 76 men. The University of North Carolina at Charlotte (UNC-C) is a branch of the state university system, nominally integrated but preponderantly white, currently operated on a nonresidential basis. Due to unanticipated difficulties in collecting data, the UNC-C sample consists of subjects tested during the spring semester and also during the summer session, plus subjects from Catawba College nearby, thought to be comparable regarding SAT scores. Information was not available, but SAT scores were expected to be around 500 for both schools. Data for 38 females were collected during the UNC-C spring semester, for 30 females during the summer session, and for 4 further females during Catawba's summer session,

a total of 72. UNC-C's spring semester yielded 17 males, the summer session 25 more, and Catawba's summer session the final 27 (total 69). Students at both institutions were mostly southern, but the proportion from northern states appeared much higher for Catawba than for UNC-C.

The third school, Duke University, is also integrated, but predominantly white. Duke is private, with about 4000 undergraduates. It draws as heavily on the North for students as on the South, and entrance requirements, as largely defined by SAT scores, are high, Freshman averages being generally around 625. Seventy-six women and 75 men from Duke participated.

Subjects at Smith and UNC-C (and Catawba) were paid $1.00 for their services. Ss at Duke were given the choice of $1.00 or one hour credit toward the six required in their Introductory Psychology course. Virtually everyone chose the hour credit.

Measures and Procedure

Three tests were administered in the experimental session, the Shipley-Hartford Test of Vocabulary and Abstract Thinking (SH), Schroder's Paragraph Completion Test (PCT), and Harvey's "This I Believe" Test (TIB). The PCT and TIB were combined so as to appear to be a single task. Additionally, Ss signed a form giving E permission to obtain SAT

scores from the school's records, and gave such information as age, class in college, and home address. SAT's proved to be unavailable for many Ss, especially those in summer school. Not surprisingly, the highest proportion of SAT scores was for the Duke samples.

Actual testing occupied something under an hour. Number of Ss per session varied between 3 and about 35. At the conclusion of the testing, an explanation was offered those who were interested, and in some cases these conversations were prolonged. All tests were timed. The SH was administered first, 5 minutes being given for the 40 multiple-choice vocabulary items, then 10 minutes for the 20 abstraction items. Instructions for the former are standard for the type test involved. Items for the test of ability to abstract consist of a series of numbers, letters or words arranged according to an organizing principle. S is to deduce the principle and complete the last item in the sequence. (In the item "tam tan rib rid rat raw hip ___," for example, the correct response is hit.)

The PCT-TIB combination followed. Instructions and items appear in the Appendix. The respective sets of items appeared together, and the sequence in which they were offered was alternated. One item appeared on each page, and each item was timed. In the present case, the time limit was 130 seconds, which is toward the upper end of the range of limits suggested by Schroder and Harvey. There were 16 items; 10 from the TIB, 6 from the PCT (the first of which is a warm-up, nonscored item). The

PCT items are fairly well standardized. More variation exists in the selection of TIB items.

Scoring

The SH is scored against a key and produces four measures: vocabulary (SH-V), abstraction (SH-A), total (SH-T, the sum of SH-V and SH-A), and the "Conceptual Quotient" (SH-CQ) which is the ratio between SH-V and SH-A (CQ = (SH-A/SH-V) X 100). The CQ reflects the original intention that the test be used as a rough screening device for detecting the type of cognitive impairment likely to be experienced by mental patients, based on the assumption that vocabulary, a function of memory, will remain relatively stable, while the active skill involved in abstracting will be an early loss. Thus, a low CQ results from a relatively high SH-V score and may indicate a loss in abstract functioning. To the extent that CQ may be related to analytic thinking, it is of interest in the present case because of the possibility, based on Gardiner and Vannoy, of a relationship between this kind of thinking and complexity.

The PCT's were scored by Schroder's research assistants at his laboratory in Princeton. Schroder reports (Schroder et al., 1967) that reliabilities for his raters range between .80 and .95. Each item was awarded a score from 1 to 7 as described previously. The score used in the

analysis was the sum of the two highest item scores, which could range between 2 and 14. In practice, out of the total 440 subjects, no item received a score above 4, and no final score exceeded 7. Some items are unscorable, in which case they are given a zero. If two or more zeros occur in the protocol, no final score is given. This occurred on approximately 6% of the protocols. Two sets of PCT scores were used, the first derived as just described. The second set may be termed the Forced PCT (PCT-F) scores. In this case, subjects who had previously received zero scores were given scores based on whatever number of items were available, these scores being, again, the sum of the two highest item scores.

Two previously inexperienced raters proceeded with the TIB's as follows: Basing procedures and criteria on mimeographed material sent from Colorado by Harvey, approximately 90 of the protocols at hand were used for training purposes. The number was so large because it proved exceedingly difficult to establish reliable criteria. The protocols involved were then rerated at a later time. After training, the judges independently rated a sample of 107 protocols. Of this number, one or the other or both judges gave zeros to 30 Ss, a 28% unscorable rate, comparable to Harvey's. A contingency table approximation of chi-square for agreeing or not agreeing to score or not score protocols produced a value of 15.95, significant at well beyond the .001 level, suggesting that there was agreement between the judges as to which subjects were and were not scorable. The correlation between the judges on the remaining 77 Ss was .83. Since this was

regarded as an adequate indication of reliability, one judge rated all the remaining protocols, and her ratings were used in the analysis.

Throughout this process there was one major and intentional deviation from Harvey's recommendations. He advises that a single global score be given the subject instead of individual item scores. Since item scores do not preclude global scores, and since they would seem to permit greater flexibility in scoring as well as tighter control over biases, item scores were assigned throughout. Global scores, then, were based on the frequency of the item scores, ameliorated if it seemed appropriate, but as seldom as possible, by the judgment of the rater. This was the Global (TIB-G) TIB score.

Three further scores were based on the TIB. First, the items in the test are rather heterogeneous in nature. In spite of the fact that Harvey seems to feel that he is measuring interpersonal complexity, the relevancy of some items to interpersonal matters is not always obvious. Accordingly, the scores on the 5 items which seemed most immediately interpersonal (marked with an asterisk in the Appendix) were considered separately, and an Interpersonal (I) score given on the basis of the item scores. As with G, of course, some of these were unscorable.

The two remaining scores are forced versions of G and I. That is, wherever a zero value was given, the judge was required to attempt to determine a single rating in addition. In general, this proved possible (although distasteful), but more difficult for the I than the G scores.

These scores, then, plus SAT scores from the college records offices for those students for whom they were on file, and the information concerning age, college class, and home (residence) address, were used in the analyses to follow.

Chapter V

RESULTS

The fact that n's are typically uneven makes especially relevant the question of homogeneity of variance. Although the F_{max} test for homogeneity is intended for equal n's, Winer (1962) states that F_{max} may be applied to unequal n's if one uses the n of the largest sample in the analysis. In the case of nonhomogeneous variances plus unequal n's, the rule of thumb (Box, 1954a, 1954b) seems to be that, if the larger variances belong to the larger sample, the probability levels for the F-ratios will be conservative, whereas if the larger variances are based on the smaller samples, the probabilities will be overstated. The size of the samples involved is not so important as the ratios of their sizes. Thus, approximately equal n's appear to retain some robustness. However, when large differences in sample variances exist along with unequal n's, relationships become rather complex and "true" probabilities extremely difficult to come by. Unfortunately, this situation exists in some of the preliminary analyses below. Where important it will be noted. The two-way analyses of variance have been computed using the least squares procedure for unequal

n's outlined by Winer (1962, p. 291 ff.) where one factor has only two levels.

Means and standard deviations for the variables are presented in Table 1. Protocols were 94% scorable for the PCT, 85% for the TIB-G, and 81% for the TIB-I. (The "unscorable" rate for the TIB is lower here than Harvey reports, and lower than in the preliminary reliability check here. This is because scores for only one rater are used. "Unscorable" decisions, although significantly related, do not overlap perfectly. Thus, the more raters, the more rejects.) For the PCT, the percentages of scores from 2 through 7 were, respectively, 3, 9, 39, 28, 12, and 3. For the TIB-G, percentages for scores 1 to 4 were 36, 15, 26, and 8, and for the TIB-I, 31, 12, 31, and 7.

The various forced scales correlated perfectly with their corresponding nonforced values, since the only discrepancies between the members of a pair occur when an S received a zero score in the nonforced case, and these are removed from the correlation. Since four scores are derived from the TIB, however, four pairings may deviate from 1.0. These were calculated separately for each of the six samples and also for the combined sample (Table 2).[1] The range for these 28 correlations was from .71 to .87, with 22 of them equalling or exceeding .80. In short,

---

[1]These and the other analyses herein reported were run at the Duke University Computer Center, using programs written by the author, and funds provided by the University.

Table 1

Means, Standard Deviations and N's[a]

		Shipley-Hartford				
		Age	Vocab.	Abs.	Total	CQ
Smith Females	$\bar{X}$	19.5	28.2	29.2	57.4	99.0
	s	1.15	3.50	5.11	7.42	10.6
	n	72	70	70	70	70
Smith Males	$\bar{X}$	19.8	27.5	29.3	56.8	100.9
	s	1.90	3.47	6.00	7.98	13.1
	n	76	76	76	76	76
UNC-C Females	$\bar{X}$	22.6	31.5	33.6	65.1	102.2
	s	6.79	3.32	3.57	4.79	10.7
	n	72	72	72	72	72
UNC-C Males	$\bar{X}$	21.4	31.6	33.3	64.9	100.8
	s	3.84	3.77	4.77	7.32	10.0
	n	69	69	69	69	69
Duke Females	$\bar{X}$	18.7	34.2	36.3	70.6	102.1
	s	2.46	2.49	2.64	3.69	7.87
	n	74	76	76	76	76
Duke Males	$\bar{X}$	18.7	34.3	36.5	70.8	102.0
	s	1.18	2.28	2.86	4.10	6.88
	n	75	75	75	75	75
Combined	$\bar{X}$	20.1	31.3	33.1	64.3	101.2
	s	3.74	4.12	5.22	8.27	10.1
	n	438	438	438	438	438

[a]N's vary because of missing data or unscorable protocols.

Abbreviations:

SH-V	Shipley-Hartford Vocabulary
SH-A	" " Abstraction
SH-T	" " Total (V + A)
SH-CQ	" " Conceptual Quotient ((A ÷ V) x 100)
SAT-V	Scholastic Aptitude Test--Verbal
SAT-M	" " " --Mathematical
PCT	Paragraph Completion Test--Score
PCT-F	" " " --Forced Score
TIB-G	"This I Believe" Test--Global Score
TIB-GF	" " " " --Forced Global Score
TIB-I	" " " " --Interpersonal Score
TIB-IF	" " " " --Forced Interpersonal Score

Table 1 (continued)

SAT		PCT		TIB			
V	M	PCT-F	PCT	G-F	G	I-F	I
386	372	4.15	4.16	1.94	1.98	2.22	2.29
83	75	.85	.86	.98	.99	.99	.99
54	54	72	68	71	55	60	56
374	400	4.34	4.37	1.86	1.73	1.75	1.75
86	83	.95	.96	.88	.85	.87	.89
48	48	76	68	76	59	64	55
474	477	4.50	4.52	1.86	1.78	1.96	1.98
71	81	.89	.88	1.05	1.03	1.06	1.06
50	50	72	67	72	64	71	62
468	509	4.41	4.36	1.88	1.97	2.08	2.09
97	80	1.19	1.20	1.01	1.02	1.03	1.06
49	49	69	66	69	59	64	53
614	621	4.58	4.63	2.40	2.40	2.50	2.52
88	74	1.06	1.05	1.10	1.11	1.09	1.11
75	75	76	71	76	70	71	66
615	663	4.88	4.88	2.35	2.39	2.39	2.39
74	73	1.16	1.15	1.01	1.04	.92	.93
71	71	75	72	75	66	70	65
505	524	4.48	4.49	2.05	2.06	2.15	2.18
130	134	1.04	1.04	1.03	1.04	1.03	1.04
347	347	440	412	439	373	400	357

Abbreviations:

PCT	Paragraph Completion Test--Score
PCT-F	" " " --Forced Score
TIB-G	"This I Believe" Test--Global Score
TIB-GF	" " " " --Forced Global Score
TIB-I	" " " " --Interpersonal Score
TIB-IF	" " " " --Forced Interpersonal Score

Table 2

Complexity Score Intercorrelations and N's[a]

		Smith Females					Smith Males				
		2	3	4	5	6	2	3	4	5	6
1. PCT-F	r	1.00	.10	.16	.00	.01	1.00	.04	-.07	.11	.07
	n	68	71	55	60	56	68	76	59	64	55
2. PCT	r		.07	.12	-.04	-.04		-.05	-.07	.10	.04
	n		67	52	56	52		68	53	57	49
3. TIB-GF	r			1.00	.80***	.79***			1.00	.86***	.84***
	n			55	60	56			59	64	55
4. TIB-G	r				.85***	.84***				.84***	.83***
	n				49	45				56	51
5. TIB-IF	r					1.00					1.00
	n					56					55
6. TIB-I	r										
	n										

[a]N's vary because of unscorable protocols.

***$p < .001$.

Table 2 (continued)

		UNC-C Females					UNC-C Males				
		2	3	4	5	6	2	3	4	5	6
1. PCT-F	r	1.00	.14	.12	.13	.18	1.00	.08	.12	.13	.17
	n	67	72	64	71	62	66	69	59	64	53
2. PCT	r		.16	.15	.15	.21		.09	.13	.09	.14
	n		67	59	66	57		66	57	61	51
3. TIB-GF	r			1.00	.79***	.81***			1.00	.71***	.74***
	n			64	71	62			59	64	53
4. TIB-G	r				.79***	.82***				.85***	.83***
	n				64	58				56	49
5. TIB-IF	r					1.00					1.00
	n					62					53
6. TIB-I	r										
	n										

Table 2 (continued)

		Duke Females					Duke Males				
		2	3	4	5	6	2	3	4	5	6
1. PCT-F	r	1.00	-.13	-.09	-.16	-.11	1.00	.01	.07	.18	.17
	n	71	76	70	71	66	72	75	66	70	65
2. PCT	r		-.10	-.05	-.09	-.03		.00	.06	.17	.17
	n		71	65	66	61		72	63	67	62
3. TIB-GF	r			1.00	.80***	.82***			1.00	.80***	.79***
	n			70	71	66			66	70	65
4. TIB-G	r				.82***	.83***				.87***	.87***
	n				66	61				62	58
5. TIB-IF	r					1.00					1.00
	n					66					65
6. TIB-I	r										
	n										

Table 2 (continued)

		Combined				
		2	3	4	5	6
1. PCT-F	r	1.00	.05	.08	.08	.09
	n	412	439	373	400	357
2. PCT	r		.06	.09	.08	.10
	n		411	349	373	332
3. TIB-GF	r			1.00	.80***	.81***
	n			373	400	357
4. TIB-G	r				.84***	.85***
	n				353	322
5. TIB-IF	r					1.00
	n					357
6. TIB-I	r					
	n					

they are highly comparable. Parallel analyses were run on all six scores in nearly all cases, but with rare exceptions, parallel results were found for scores deriving from the same test. Accordingly, unless the others are of special interest, only the results for the PCT and TIB-G scores (i.e. the scores as used by their respective creators) will be considered. All t tests and correlations are two-tailed.

Preliminary Tests

The initial question asked of the data was whether the sequence in which the Ss took the PCT and TIB affected their scores. Ss were divided according to which test they had taken first, and the two groups compared on each of the complexity scores by t tests. These were uniformly non-significant. The sequence in which the PCT and TIB were taken did not affect the scores (see Table 3).

The second issue was whether the UNC-C sample, drawn as it was from several sources, was homogeneous enough to be regarded as a single sample. Since the males were relatively evenly distributed among the three sources (n's were 17, 25, 27), one-way analyses of variance were calculated on the complexity scores and on each of the six intelligence scores (Table 4). These tests were all nonsignificant, so there is no obstacle to regarding the UNC-C males as representing a single population.

Abbreviations:

PCT	Paragraph Completion Test--Score
PCT-F	" " " --Forced Score
TIB-G	"This I Believe" Test--Global Score
TIB-GF	" " " " --Forced Global Score
TIB-I	" " " " --Interpersonal Score
TIB-IF	" " " " --Forced Interpersonal Score

Table 3

Preliminary Tests

		Complexity Scores					
		PCT-F	PCT	TIB-GF	TIB-G	TIB-IF	TIB
A. Effects on complexity scores of the sequence in which tests are taken	t	.04	.04	.52	.88	.93	.39
	d.f.	438	410	437	371	398	355
B. UNC-C students 23 years old or older v. younger UNC-C students	t	1.86	1.64	-.23	-.05	-.29	-.61
	d.f.	139	131	139	121	133	113
	F-ratios						
C. UNC-C College Class x Sex two-way ANOVA	Class	8.40**	6.30*	4.98*	4.10*	2.96	2.99
	Sex	1.00	1.02	.09	.41	.07	.02
	C x S	.05	.00	2.90	1.83	4.90*	3.03
	d.f.	1, 121	1, 115	1, 121	1, 103	1, 115	1, 95
D. J. C. Smith Univ. College Class x Sex two-way ANOVA	Class	1.93	1.88	.06	.01	.01	.17
	Sex	1.92	1.85	.35	2.11	7.59**	8.59**
	C x S	.08	.15	.01	.07	.57	.43
	d.f.	1, 144	1, 132	1, 143	1, 110	1, 120	1, 107
E. Duke University Residence x Sex two-way ANOVA	Resid.	.01	.01	1.63	.88	.38	.15
	Sex	2.70	1.71	.04	.01	.33	.48
	R x S	.24	.01	.83	1.11	2.78	1.71
	d.f.	1, 147	1, 139	1, 147	1, 132	1, 137	1, 127

*$p \leq .05$; **$p \leq .01$

Abbreviations:

SH-V	Shipley-Hartford Vocabulary
SH-A	" " Abstraction
SH-T	" " Total (V + A)
SH-CQ	" " Conceptual Quotient ((A ÷ V) x 100)
SAT-V	Scholastic Aptitude Test--Verbal
SAT-M	" " " --Mathematical
PCT	Paragraph Completion Test--Score
PCT-F	" " " --Forced Score
TIB-G	"This I Believe" Test--Global Score
TIB-GF	" " " " --Forced Global Score
TIB-I	" " " " --Interpersonal Score
TIB-IF	" " " " --Forced Interpersonal Score

Table 4

UNC-C Sample Homogeneity

		Complexity Scores					
A. UNC-C males		PCT-F	PCT	TIB-GF	TIB-G	TIB-IF	TIB-I
spring v. summer v. Catawba subsamples one-way ANOVA	F	2.70	2.18	1.15	1.18	1.39	2.06
	d.f.	2, 66	2, 63	2, 66	2, 56	2, 61	2, 50
		Intelligence Scores					
		SH-Voc.	SH-Abs.	SH-Tot.	SH-CQ	SAT-Verb.	SAT-Math.
	F	2.09	.61	.83	2.11	.46	2.14
	d.f.	2, 66	2, 66	2, 66	2, 66	2, 46	2, 46
		Complexity Scores					
B. UNC-C females		PCT-F	PCT	TIB-GF	TIB-G	TIB-IF	TIB-I
spring v. summer subsamples t tests	t	-1.89	-2.56*	1.65	1.90	1.97	1.59
	d.f.	70	65	70	62	69	60
		Intelligence Scores					
		SH-Voc.	SH-Abs.	SH-Tot.	SH-CQ	SAT-Verb.	SAT-Math.
	t	-2.75**	.25	-1.66	1.69	-.07	.56
	d.f.	70	70	70	70	48	48

*$p \leq .05$

**$p \leq .01$

F_{max} is significantly large for the three groups in the analysis of SAT-V scores, and the larger of the variances is based on the smaller n. Even though this combination is likely to lead to an unrealistically improbable result (Box, 1954a), the ANOVA F is nonsignificant.

Since only four female Ss were from Catawba, these were combined with the UNC-C female summer session Ss, and t tests were used to compare this group with the UNC-C female spring semester Ss. In this case the difference between the two groups on their PCT scores was significant at a probability level between .01 and .02, the summer session women having the higher average (Table 4). The difference between the two groups on their TIB-G scores has an associated probability between .06 and .07, which is generally considered nonsignificant, but is of interest here for two reasons: First, the two groups had different variances on TIB-G, and since the larger variance was based on the larger n, the probability of the difference may be slightly conservative (the n's are not greatly different). Second, the direction of the difference is opposite that on the PCT, i.e. the spring semester students have the higher average.

The same two groups were compared on the intelligence measures. The two groups differed significantly ($.01 < p < .02$) only on the SH-V measure, the difference favoring the summer school students.

Thus, while the UNC-C male sample may be treated as a single unit, there is reason to doubt the homogeneity of the female group. Analyses in which this might prove important were generally run three times, once

with the combined UNC-C female sample, and once with each of the subgroups.

Complexity, Age, College Class, Sex and Residence

Since complexity is often discussed in developmental terms, and since roughly 20% of the UNC-C sample (combining males and females) were older than normal for undergraduates, a series of *t* tests were run on complexity scores to check for age effects. In this situation, 23 or over was "old" and under 23 "young." These tests were all nonsignificant (Table 3, p. 63).

As a second rough check for age effects (rough since clearly in this case, as well as the one above, selection factors other than age are also operating) Sex x Class-in-College two-way analyses were run separately for the Smith and UNC-C samples, class being considered lower (Freshman and Sophomore) vs. upper (Junior and Senior). No such analysis was performed on the Duke sample because there were too few upperclassmen.

For the UNC-C sample, a class effect existed such that upperclassmen were more complex than lower on four of the six complexity measures, including the PCT and TIB-G (Table 3). F_{max} is nearly significant at the .05 level for the PCT analysis, but again large variances are based on large n's, so the results are probably conservative. In this case, older

students, as defined by college experience, are more complex than younger, a finding perhaps supportive of a developmental position. (Apropos the previous finding of nonsimilarity between the two female groups, the spring Ss had 9 upperclassmen, the summer group 13. The difference, even expressed as proportions, is not large. Such as it is, the difference is consistent with the pair of results for the PCT and contradictory for the TIB-G findings.)

The college class difference was not replicated in the Smith sample, however (Table 3, p. 63). In this case, there were no class effects whatsoever. This analysis, however, constitutes the sole instance in which the TIB-G and TIB-I provided results different from each other in an interesting way. The sex effects are highly significant for the TIB-I and not at all so for the TIB-G, this despite their correlations of .83 and .84 for the two samples from Smith. The difference between the sexes, females having the higher average, seems to result from the fact that a fairly large proportion of the Smith males express severe doubts about the permanence and reliability of friendship, such distrust and cynicism leading to scores of 1 or 2 in Harvey's system. This finding did not appear in the UNC-C Class x Sex analysis (nor in the Duke ANOVA, below), so it represents either a minority group phenomenon (not an implausible hypothesis) or sampling error.

Duke's selection practices permit a test of whether residence in the South or North affects complexity scores. The question is relevant because

Southern education is often regarded as inferior. Accordingly, a Residence x Sex ANOVA was run. Neither sex nor residence produced significant results (Table 3, p. 63). Apparently, either residence is unimportant, or becomes so when intelligence is high. (No check was made on the comparability of intelligence scores for the two residence groups. For the Duke women, several of the intelligence scores did correlate significantly with the TIB-G.)

Complexity, Schools and Intelligence

College x Sex ANOVA's were run for both intelligence and complexity scores. Since the heterogeneity of the UNC-C female (UF) sample might be important here, each analysis was done three times, once each with the combined UF sample, the UF-spring subsample, and the UF-summer subsample. For the combined UF sample, n's in the analyses tend to be approximately equal. The subsamples make the n's fairly unequal. There is no serious problem for the complexity ANOVA's because variances are homogeneous. For the UF-spring subsample, F_{max} is borderline significant for the PCT, but the large variance and large n's go together, mitigating the problem.

The situation is otherwise for the intelligence score comparisons. For all three UF samples all variances based on the Shipley-Hartford are

significantly (or nearly so) different, and in at least one of the cases SAT-V is also nonhomogeneous. For two of the variations either sample sizes are roughly equal (combined UF), or there is a favorable combination of sample size and variance magnitude (UF-spring). Not even that much can be said for the UF-summer sample. However, the pattern of ANOVA results is identical in all three cases, and is in accord with expectations and selection criteria: There is an enormous intelligence effect across schools. F's range from approximately 100 to approximately 350. The only intelligence-based measure not related to schools is the SH Conceptual Quotient (Abstraction divided by Vocabulary and multiplied by 100). Additionally, there is a sex main effect on SAT-M, males, as would be expected, being higher (Table 5).

The results for complexity are similar but not so overwhelming (Table 6). All complexity scores for all three analyses show significant school effects. For the combined UF sample, only the school effects are significant. For the subsamples, additional significant results occur in a pattern which follows logically from the differences in their scores: When the UF-spring group was used (who were lower on the PCT and higher on the TIB than were the UF-summer Ss), significant sex effects appear, with males averaging higher on the PCT and females higher on the TIB. With the UF-summer group, sex effects disappear, and are replaced with interactions: On the PCT, Smith males, UNC-C females and Duke males have higher means than the opposite sexes at their respective schools. The

Table 5

Sex x School

Two-way ANOVA of intelligence scores separately for UNC-C female sample conditions

		Intelligence Score F-Ratios with Degrees of Freedom					
	Source	SH-V	SH-A	SH-T	SH-CQ	SAT-V	SAT-M
With complete UNC-C female sample	Sex	.27	.01	.13	.01	.22	17.71***
	d.f.	1,432	1,432	1,432	1,432	1,341	1,341
	Schools	152.26***	102.56***	184.31***	1.64	247.98***	341.19***
	S x S	.84	.10	.20	.99	.17	.29
	d.f.	2,432	2,432	2,432	2,432	2,341	2,341
With spring UNC-C female sample	Sex	.02	.01	.01	.05	.19	16.04***
	d.f.	1,398	1,398	1,398	1,398	1,328	1,328
	Schools	155.70***	100.70***	179.59***	1.99	242.19***	336.81***
	S x S	2.41	.12	.37	2.12	.16	.35
	d.f.	2,398	2,398	2,398	2,398	2,328	2,328

Table 5 (continued)

With summer UNC-C female sample	Sex	1.96	.01	.43	.69	.18	17.07***
	d.f.	1,394	1,394	1,394	1,394	1,304	1,304
	Schools	155.37***	98.10***	176.88***	1.68	238.33***	346.25***
	S x S	1.12	.04	.41	.40	.17	.29
	d.f.	2,394	2,394	2,394	2,394	2,304	2,304

***$p \leq .001$

Abbreviations:

SH-V	Shipley-Hartford Vocabulary
SH-A	" " Abstraction
SH-T	" " Total (V + A)
SH-CQ	" " Conceptual Quotient ((A $\div$ V) x 100)
SAT-V	Scholastic Aptitude Test--Verbal
SAT-M	" " " --Mathematical

Table 6

Sex x School

Two-way ANOVA of complexity scores separately for UNC-C female sample conditions

Samples	Source	Complexity Score F-Ratios with Degrees of Freedom					
		PCT-F	PCT	TIB-GF	TIB-G	TIB-IF	TIB-I
With complete UNC-C female sample	Sex	1.96	.99	.16	.03	2.01	2.85
	d.f.	1,434	1,406	1,433	1,367	1,394	1,351
	Schools	8.28***	8.22***	11.55***	12.12***	9.03***	7.39***
	S x S	1.43	1.58	.11	1.39	2.80	2.98
	d.f.	2,434	2,406	2,433	2,367	2,394	2,351
With spring UNC-C female sample	Sex	4.03*	3.16	.83	.75	4.76*	5.25*
	d.f.	1,400	1,375	1,399	1,338	1,361	1,323
	Schools	8.97***	9.46***	9.28***	9.52***	7.13***	5.73**
	S x S	.32	.17	.11	.51	1.31	1.64
	d.f.	2,400	2,375	2,399	2,338	2,361	2,323

Table 6 (continued)

With summer UNC-C female sample	Sex	1.26	.40	.01	.02	1.17	2.15
	d.f.	1,396	1,370	1,395	1,332	1,356	1,317
	Schools	7.97***	7.78***	12.76***	12.87***	9.69***	7.57***
	S x S	2.51	3.35*	.85	2.89	4.78**	4.40*
	d.f.	2,396	2,370	2,395	2,332	2,356	2,317

*$p \leq .05$

**$p \leq .01$

***$p \leq .001$

Abbreviations:

PCT	Paragraph Completion Test--Score
PCT-F	" " " --Forced Score
TIB-G	"This I Believe" Test--Global Score
TIB-GF	" " " " --Forced Global Score
TIB-I	" " " " --Interpersonal Score
TIB-IF	" " " " --Forced Interpersonal Score

direction of the interaction is reversed on the TIB, and does not quite reach significance for the TIB-G scores. For TIB-I, however, Smith females, UNC-C males and Duke females have the higher means, and the interaction is significant.

All three complexity scores (PCT, TIB-G, TIB-I) provided significant school effects, and in each case the mean for Smith was lowest and that for Duke highest. For the two TIB scores the differences between Smith and UNC-C were extremely small; but on the PCT, UNC-C is more nearly in the middle. In view of the difference between Smith and UNC-C on SAT scores (approximately 100 points) and race, with all the implications associated with the latter, it is of particular interest to know which groups differed from each other. Accordingly, for each complexity score individual comparisons (Winer, 1962) were made for all pairs of schools in the ANOVA using the combined UNC-C female sample. The pattern of results is precisely the same for all three scores: Smith and UNC-C are not different from each other, but both are different from Duke. This pattern holds whether one selects critical values according to "risky" or "conservative" principles.

PCT, TIB and Intelligence

Results regarding the relationships between the PCT and TIB (see

Table 2, p. 61) may be succinctly stated: The two tests yield orthogonal scores. No significant correlations were found for either the PCT and TIB-G or the PCT and TIB-I, for any of the six samples or for the entire combined sample. The correlations ranged between -.07 and +.21.

Nor was there support for the earlier suggestion that Harvey's Systems I and II, and probably III, were discriminations of Schroder's simple scores, whereas Harvey's IV's were like Schroder's 6's (and above). A chi-square testing this hypothesis (TIB 1-3 vs. 4, PCT less than 6 vs. 6 and 7) was nonsignificant.

Both the PCT and TIB-G correlate more strongly with intelligence than they do with each other (Tables 7 and 8). Further, the pattern of correlations with the intelligence measures is basically the same for both tests. In my opinion, there is no great cause for concern, however. The proportion of variance involved in these correlations is never very large and is typically quite small (though statistically significant).

The most important set of these correlations, of course, is for the combined samples. This provides a particularly broad range of intelligence scores, and an N which typically ranges between 300 and 400. In this case, we find that both the PCT and the TIB-G are significantly ($p < .01$) correlated with SH-Vocabulary, SH-Abstraction, SH-Total, SAT-Verbal and SAT-Mathematics. (It should be noted that SH-T is positively correlated with SH-V and SH-A.) In magnitude the correlations range from .14 to .27, so at the most about 8% of the variance is involved. A few of the correla-

Table 7

Correlations Between Intelligence Measures and the Paragraph Completion Test (PCT)

		Shipley-Hartford				SAT	
		Vocab.	Abs.	Total	CQ	Verbal	Math.
Smith Females	r	.23	.19	.24	.02	.02	.08
	n	66	66	66	66	50	50
Smith Males	r	.19	.09	.15	-.04	.07	.07
	n	68	68	68	68	46	46
UNC-C Females	r	.30*	.01	.22	-.18	.10	.03
	n	67	67	67	67	47	47
UNC-C Males	r	.18	.08	.14	-.05	.06	-.01
	n	66	66	66	66	47	47
Duke Females	r	-.01	-.05	-.04	-.04	.04	.14
	n	71	71	71	71	70	70
Duke Males	r	.18	.05	.13	-.07	.12	.16
	n	72	72	72	72	68	68
Combined Samples	r	.25***	.16**	.23***	-.04	.18**	.20***
	n	410	410	410	410	328	328

*$p \leq .05$

**$p \leq .01$

***$p \leq .001$

Abbreviations:

SH-V	Shipley-Hartford Vocabulary
SH-A	" " Abstraction
SH-T	" " Total (V + A)
SH-CQ	" " Conceptual Quotient ((A ÷ V) x 100)
SAT-V	Scholastic Aptitude Test--Verbal
SAT-M	" " " --Mathematical

Table 8

Correlations Between Intelligence Measures and the "This I Believe" Test (TIB)

		Shipley-Hartford				SAT	
		Vocab.	Abs.	Total	CQ	Verbal	Math
Smith Females	r	.07	-.03	.01	-.08	.17	.13
	n	53	53	53	53	41	41
Smith Males	r	-.06	.33*	.22	.40**	.10	.26
	n	59	59	59	59	38	38
UNC-C Females	r	.26*	.08	.23	-.15	.17	.26
	n	64	64	64	64	44	44
UNC-C Males	r	.19	-.02	.09	-.15	.10	.08
	n	59	59	59	59	41	41
Duke Females	r	.23	-.24*	-.02	-.35**	.24*	.02
	n	70	70	70	70	69	69
Duke Males	r	.11	.01	.07	-.07	.23	.03
	n	66	66	66	66	62	62
Combined Samples	r	.23***	.14**	.20***	-.05	.27***	.23***
	n	371	371	371	371	295	295

*$p \leq .05$

**$p \leq .01$

***$p \leq .001$

Abbreviations:

SH-V	Shipley-Hartford Vocabulary
SH-A	" " Abstraction
SH-T	" " Total (V + A)
SH-CQ	" " Conceptual Quotient ((A ÷ V) x 100)
SAT-V	Scholastic Aptitude Test--Verbal
SAT-M	" " --Mathematical

tions for the individual samples are larger, but even the largest is only .40.

Within the samples no impelling pattern seems to exist. In two cases, Smith males and Duke females, SH-Abstraction and SH-Conceptual Quotient correlated significantly with TIB-G. (Again, SH-A and SH-CQ are not independent.) However, for the latter group the two correlations are negative, while in the former they are positive. This ambiguity, plus the weak relationship between SH-A and the complexity scores, makes unlikely my earlier suggestion that complexity is based on analytic thinking, at least insofar as SH-A or SH-CQ measure this skill.

This weak relationship is consistent with Schroder's prediction regarding his PCT and the Abstraction score. His expectation was based on two considerations: The SH-A tasks are not interpersonal and they are not complex enough to fully engage complex thinking. On the other hand, his prediction of nonrelationship between the PCT and vocabulary and/or verbal scores is contradicted by the present findings. That is, Schroder argues that if anything the PCT will be related to complex information-processing or decision-making aspects of intelligence, but not to verbal skills per se, and particularly not to memory skills (e.g. vocabulary). In this study, both the PCT and the TIB are related to vocabulary and to verbal skills. This finding is consistent with Harvey's report (1966).

Finally, since Harvey typically analyzes his data using group means instead of correlations, in each of the six samples Ss were grouped by each complexity score (separately) and one-way ANOVA's run for each of the

intelligence scores and for age (four groups were used for the PCT also: scores of 2-3, 4, 5, 6-7). No surprising results were discovered. In a few cases (four), significant correlations between complexity and intelligence had been found, but the present F ratios were nonsignificant. This is presumably due to greater power on the part of the correlation in testing a linear hypothesis, plus its lack of concern for unequal n's, etc. In three instances the correlations and F's were both significant. There were six significant F/nonsignificant r combinations. Three of these were for groups as defined by TIB-I scores, for which correlations had not been calculated. No particular pattern appears for these results, except that on the TIB there is some tendency for Levels II and IV to be alike and I and III to be alike. This latter pattern, but not the linear findings, are consistent with Harvey (1966).

Reliability

Of course, the best way to learn about the test-retest reliability of these tests would be to devise alternate forms and administer both versions to some Ss. Lacking this, an estimate of the reliability, coefficient alpha (Rozeboom, 1966), can be calculated on the basis of the variances of the items and of the total scores. The scoring procedures used above, however, do not lead to values which can be used to calculate alpha. Accord-

ingly, the PCT and TIB item scores were used to derive total scores compatible with the application of alpha. In each case, the average of the scored items was found, and this value multiplied by the total number of items (5 and 10, respectively). The variance of these scores was used in finding the alpha coefficients.

In general, these scores (SumPCT and SumTIB) behaved as the other scores. Summed scores generally correlated in the .80's with the corresponding regular scores (SumTIB with TIB-I being mostly in the .70's). Also, the summed scores tended not to correlate significantly with scores based on the other test. Here, however, there were exceptions. For the Smith females, SumPCT correlated .235 with SumTIB and .241 with TIB-G, both significant at the .05 level; and in the Duke male sample, SumPCT barely misses significance with TIB-I ($r = .231$). With the combined samples, none of the intertest correlations approach statistical significance.

For the combined samples, the estimates of test-retest reliability are .406 for the PCT and .776 for the TIB. Within the samples, the values for the PCT range between .294 and .515, and between .605 and .814 for the TIB (Table 9). On these grounds the TIB can be regarded as reliable, and the PCT cannot.

Table 9

Alpha Coefficients

	Samples						
	Smith Females	Smith Males	UNC-C Females	UNC-C Males	Duke Females	Duke Males	Combined
Alpha for PCT	.29	.44	.28	.52	.31	.43	.41
Alpha for TIB	.65	.61	.76	.72	.81	.77	.78

Chapter VI

DISCUSSION

Summary of Results

The major purposes of this study were outlined at the close of Chapter III. The relevant results are as follows:

A. The main point of interest seems to have been unequivocally resolved: Schroder's PCT and Harvey's TIB are uncorrelated with each other, and are therefore not measuring the same variable. Even when the TIB scoring is revised to make it more obviously interpersonal in orientation (and it should be emphasized that even the global score is assumed to be interpersonal), no relationship with the PCT exists.

The current data cannot tell us whether this independence is based on the failure of one test to measure what it is supposed to, or whether one of the measures is "better" than the other. As has been shown, the scoring criteria are vastly different for the two, but each has a basis in the respective theory. The suggestion that the two theories might be dealing with

different portions of the dimension was not substantiated.

The most obvious consequence of this is that the theories and associated research are irrelevant to each other. Irrespective of similarities in background, language, terminology, etc., it is simply not appropriate to combine the findings in attempting to define or understand "integrative complexity."

B. A second major question is the relationship of complexity with intelligence. The pattern is the same for both the PCT and the TIB: complexity is significantly related to measures of simple vocabulary and abstraction ability, and to verbal and mathematical skills as measured by the SAT. These findings essentially replicate those of Harvey (1966), but are much weaker than the Schroder et al. (1967) report of correlations in the .40's when the variance of intelligence scores is great. The results contradict Schroder's prediction that complexity should not correlate with verbal skills, and especially should not correlate with memory functions (vocabulary).

The correlations with intelligence, while statistically significant, were not large, and were certainly not large enough to threaten the value or standing of either test on this basis alone, provided it is understood that either measure of complexity has more in common with intelligence than with the other complexity measure. The relationship with intelligence is strong enough that research using either measure is of questionable value unless controls for intelligence are incorporated. Schroder routinely

matches Ss for intelligence (e.g. Schroder et al., 1967); Harvey (e.g. Harvey, 1966) often does not mention it, and therefore presumably does not do it.

C. As to the existence of sex differences in the complexity scores, results are not clear-cut. Several alternative answers are available depending on which sample and which analysis one wishes to rely upon. The broadest answer, based on the Sex x Schools ANOVA with the complete UNC-C female sample, is that there are no sex effects. However, the same analysis performed separately with the two distinguishable UNC-C female subsamples did lead either to sex or interaction effects.

At least one ANOVA was done independently for each school in which sex differences could have been demonstrated. Only Smith showed a sex effect, and that was for the TIB-I score; that is, a strongly interpersonal measure based on the TIB. In this case, females were more complex, a result which makes some sense in terms of the kinds of stresses to which minority groups are likely to be exposed (e.g. the greater difficulty the male has in playing his role). Since students from Smith are probably also the lowest of the three groups socioeconomically, this finding partially replicates Gardiner's (1968), that among children from lower class backgrounds females were more complex than males. The question of sex differences, then, is still open. Whether a difference will exist apparently depends upon population characteristics. A check should probably be made in every situation in which a difference might be important.

D. Differences in complexity scores among schools could result from any of many factors, the most obvious being intelligence, race and socioeconomic background, these being all interrelated. The mean complexity scores for the three schools did fall in the expected sequence, but individual comparisons show that the differences between Smith and UNC-C are nonsignificant, and that both differ significantly from Duke. This means that neither race per se, nor fairly large intelligence differences (approximately 1 sigma on the SAT), are sufficient to indicate complexity differences. Of course, this could be attributed to a floor in the complexity scores, but that problem is intrinsic in the tests and theories.

E. Test reliability, as estimated by coefficient alpha, is unacceptably low for the PCT (.41), but reasonably high for the TIB (.78). The inter-item correlations for the PCT were lower than those for the TIB (the former ranged between .05 and .17, and the latter between .15 and .47, and were mostly in the .20's), and were much lower than those reported in Schroder et al. (1967) (range .39 to .62).

It should be kept in mind that the scores used to calculate coefficient alpha are not those regularly used in research. In the case of the PCT, the difference is probably not important: Although the score currently in favor is the sum of the two highest item scores, the mean of all the item scores has been used and is presumably still acceptable. Using the sum of the item scores (but adjusting for nonscored items) instead of the mean is necessary for calculating alpha, but does not change score relationships.

The set of scores derived from the TIB and used in finding alpha bears no such close relationship to regular TIB scores, for the simple reason that Harvey does not assign individual item scores. This practice seems to me to be a weakness in that it requires the scorer to somehow bear in mind 10 (the number of items) sources of information, which is essentially not possible. In my opinion, the kernel of this objection holds even though the rater keeps some sort of mental cumulative tally: Why not write it down? In the present case we did, separately for each item. This enabled us to obtain the Interpersonal score, and later to sum the item scores so as to estimate reliability. In both cases, the summed scores correlated highly with the regular scores.

The low reliability of the PCT is something of a threat to it, especially in view of the fact that high scores are so rare: The misclassifications which result from low reliability are a serious problem. That Schroder's research results conform so nicely to theory suggests, in light of this low reliability, that his cutoff score may be higher than would be necessary for a more reliable measure. Increasing the number of items (and reverting to mean scores) would probably help to stabilize the measure. Of course, it would be valuable to have an actual test-retest coefficient instead of an estimate, and responsibility for this is basically Schroder's, since he is the person best equipped to devise alternative items and provide authoritative ratings. Until such an effort is mounted, the reliability of the instrument may be regarded as unsatisfactory.

The situation of the TIB is different. Although coefficient alpha is acceptable, the onus for a true test-retest reliability correlation for the TIB rests on Harvey, just as it does on Schroder for the PCT, in spite of the present satisfactory value of alpha. This is so because, beyond the responsibility of the test creator to provide adequate information about his test, the set of TIB scores used to calculate alpha bear a less certain relationship to the usual TIB scores than was the case for the PCT. This in turn is partly due to the fact that confusion in the TIB scoring is typically between a score of 2 or 4 on the one hand, and 1 or 3 on the other. Thus, it is not uncommon that the switch of the score on a single item (as on a retest) would lead to a change of two points in the global score, while it would be scarcely noticeable on the sum-of-items score. In short, although the correlation between TIB-G and the summed-items scores is high, the categorization of Ss into four levels based on a retest might not be so much in agreement. I would not be surprised if alpha overestimates the reliability for the TIB, but the issue could be resolved empirically more quickly and easily than by debate.

Comment and Implications

Given a choice, I personally prefer Schroder's position to Harvey's, for two reasons. First, the theory is more elegant, in that it is simpler,

more easily applied to a broad range of cognitive domains, and permits behavioral predictions which are more clearly related to the theory. By these criteria, the theory is more powerful than Harvey's, although Harvey seems to regard his theory as relating to a greater variety of behavior than Schroder, so in this sense Harvey's may be the more powerful.

The second reason I prefer Schroder's position is that his research gives the impression of greater rigor. The most obvious case in point is the difference in approach to the scoring criteria for the respective tests. Beyond this is the looseness in what Harvey is willing to call validational evidence, plus his seeming failure to control for intelligence in these validational studies. Naturally, the best comparison of the two tests would be with regard to predictive efficacy, but this was not possible in this study. It would be difficult to compare the tests using existing literature, since there seems to be very little, if any, overlap in the actual dependent variables which have been investigated.

Although Schroder's seems the stronger position, he is not without problems. At the theoretical level he has yet to elucidate the "connectedness" concept. As indicated earlier, if integrative complexity is to be distinguished from a dimensional position, it will have to be on the basis of connectedness, yet it is never made clear what this connectedness really is, or how it is represented in the theoretical model.

The second problem is the low reliability of the PCT (as estimated by coefficient alpha). The PCT inter-item correlations found here were much lower than those reported in Schroder et al. (1967). Coefficient al-

pha will become larger as the inter-item correlations increase, but one must assume that the scores (ratings) are reliably determined if alpha is to be taken as an estimate of test-retest reliability. Thus, both inter-item and inter-rater correlations must be adequate.

It is interesting to speculate as to whether the change in the application of scoring criteria (see Chapter III), since Schroder et al. (1967) is related to the reduced item intercorrelations. Schroder's explanation for the shift in distribution of PCT scores may be interpreted as implying that the 1967 inter-item correlations were inflated by the interpretations made by the raters. These interpretations, being unsystematic, should reduce inter-rater reliabilities, but the rater reliabilities reported in Schroder et al. seem fully adequate (.80-.95). The combination of item and rater intercorrelations in Schroder et al. implies a high value of alpha and some reason to expect that alpha reflects reliability. The low inter-item correlations in the present study virtually preclude high reliability, irrespective of rater reliabilities. The former situation seems preferable. Since the changes in scoring procedure were introduced because of the difficulty in distinguishing among scores at the complex end of the dimension, and since the PCT format (two or three sentences in response to a sentence stem, with a short time limit) restricts the expression of complex, abstract concepts, the most obvious first attempt to solve the low reliability problem would be to arrange for stimuli and/or time requirements which would tend to draw (or at least permit) more complex responses.

Other issues become problems only within a context. As it stands now, unless E is willing to screen large numbers of Ss in order to find the few who are complex, and has reason to expect in advance that his efforts will be rewarded (note that no Ss with scores of 7 or higher were found at Smith and only 2 at UNC-C, a complexity rate of less than 1% for a total of nearly 300 Ss), Schroder's theory, measure and research are of limited value. If one is interested in studying naturally occurring or random groups of Ss, as most of us are, complexity (defined in Schroder's terms) will be either rare or nonexistent. This problem is compounded by the scarcity of information about middle complexity Ss. There appears to be little concern for them in Schroder's work, consequently one does not know where they stand, or where to expect them to stand, relative to the extreme groups. It is possible that changing the test as suggested above would improve the distribution of scores. Learning about the behavioral characteristics of the mid-complexity group is a matter of research. At least one, and preferably both, will be necessary before the variable will become useful in most research.

Given these difficulties, the interested researcher may wish to turn to the dimensionally oriented theories and measures. Of these, Bieri (e.g. Bieri, 1966) is the leader in terms of published research, but Scott's article mentioned earlier (Scott, 1969) shows evidence of an unusual, and laudable, degree of concern for preliminary, methodological questions. Perhaps the major weakness in Scott's study is that he, too, fails to control

for intelligence. Beyond this, there seems to be considerable promise in his work.

The decision to adopt a dimensional definition of complexity may not put one so much at odds with the "integrative complexity" theorists as might be thought at first. As mentioned earlier, TIB scores correlate with REP dimensionality (Harvey, 1966), and Schroder, at least at the theoretical level, will not become clearly different from the dimensional viewpoint unless "connectedness" can be defined so as to achieve the distinction.

The correspondence between Schroder and the dimensionalists may potentially be increased in two different, but related, ways. Schroder's complex thinking seems to involve two components; first, the ability to perceive and tolerate alternative interpretations of a situation or event, this being in a sense the first stage of complexity, and second, the development of some sort of abstract basis for relating and evaluating the alternatives. It may be that the first of these could be fruitfully conceptualized as a differentiation of evaluative judgments (the development of several kinds of "good" and "bad") and/or a reduction of the correlations between an "evaluative dimension" and other nominally nonevaluative dimensions. Either of these could lead to judgments involving relative value, various kinds of value, and simultaneous positive and negative value, as well as relatively nonevaluative judgments.

The second way to increase the relationship between integrative complexity and dimensional complexity would be to assume that the connecting

or relating component of Schroder's theory represents superordinate dimensions, and that there is therefore some hierarchical organization of dimensions. Such a conceptualization is more difficult with dimensions than it would be with categories. For example, one of the few efforts to investigate a hierarchical structure, that of Zajonc (1960), did not use dimensions. It is tempting to suggest that these hypothesized superordinate dimensions are based on the person's philosophical values, but if this is to be the case, distinctions must be made between values which are and values which are not capable of preserving alternative viewpoints. The structures would be quite different in the two cases.

It would seem, then, that the insights from the theory of integrative complexity could be applied to research on natural or random groups of Ss either by (a) improving on Schroder's measuring techniques and/or studying middle as well as extreme complexity Ss, or (b) by interpreting Schroder's position in dimensional terms, and looking for the differentiation of evaluativeness or for a hierarchical organization of dimensions. Either approach could provide valuable extensions of the theory of cognitive structure.

APPENDIX

INSTRUCTIONS, ITEMS, AND A SAMPLE PAGE FOR THE PARAGRAPH COMPLETION TEST AND THE "THIS I BELIEVE" TEST

On the following pages you will be asked to write a short paragraph on your beliefs about or opinions of several different things.

On each page you will find either the beginning of a sentence (a word or phrase followed by several dots, for example, "I like . . .") or a whole sentence beginning with "This I believe (for example, "This I believe about human nature").

If the beginning sentence is unfinished, complete it and try to add at least two further sentences on the same topic.

If the beginning sentence is complete, try to write at least three sentences on the topic.

In other words, in either case try to write at least three sentences.

These answers cannot and will not be graded right or wrong, so please write what you genuinely feel or believe. Work quickly, but be as clear as possible.

You will be given 130 seconds for each answer. After 110 seconds I

will say "Finish your sentence." Make sure your last sentence is completed. At 130 seconds I will say "Turn to the next page." You should begin on the next page immediately.

You should write on the topics in the order of their appearance. Wait to turn each page until I give the signal. Once you have turned a page, do not go back to it.

Do not turn this page until you are given the signal.

Please raise your hand if you have any questions.

Items for the Paragraph Completion Test (PCT):

When preparing for an exam --- (this is a non-scored, warmup item)

Rules---

When I am criticized---

When I am in doubt---

When others criticize me it usually means---

Confusion---

Items for the "This I Believe" Test (TIB):

This I believe about the American way of life.

*This I believe about marriage.

*This I believe about people.

This I believe about religion.

*This I believe about compromise.

*This I believe about friendship.

*This I believe about sin.

This I believe about foreign aid.

This I believe about capital punishment.

This I believe about world government.

*Items marked with an asterisk were those used to determine the Interpersonal scores (TIB-I and TIB-IF).

Sample Page

When preparing for an exam.....

Try to write at least 3 sentences.

Do not turn this page until you are given the signal.

REFERENCES

REFERENCES

Allard, M., & Carlson, E. R. The generality of cognitive complexity. Journal of Social Psychology, 1963, 59, 73-75.

Bieri, J. Cognitive complexity-simplicity and predictive behavior. Journal of Abnormal and Social Psychology, 1955, 51, 263-268.

Bieri, J. Complexity-simplicity as a personality variable in cognitive and preferential behavior. In D. W. Fiske & S. R. Maddi (Eds.), Functions of varied experience. Homewood, Ill.: The Dorsey Press, 1961. Pp. 355-379.

Bieri, J. Cognitive complexity and personality development. In O. J. Harvey (Ed.), Experience, structure and adaptability. New York: Springer, 1966. Pp. 13-37.

Bieri, J., & Blacker, E. The generality of cognitive complexity in the perception of people and inkblots. Journal of Abnormal and Social Psychology, 1956, 53, 112-117.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. Annals of Mathematical Statistics, 1954, 25, 290-302. (a)

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-498. (b)

Carr, J. E. Cognitive complexity: Construct descriptive terms vs. cognitive processes. Psychological Reports, 1965, 16, 133-134.

Clayton, M. B., & Jackson, D. N. Equivalence range, acquiescence, and overgeneralization. Educational and Psychological Measurement, 1961, 21, 371-382.

Gardiner, G. S. Some correlates of cognitive complexity. Unpublished master's thesis, University of Alberta, 1968.

Gardner, R. W. Cognitive styles in categorizing behavior. Journal of Personality, 1953, 22, 214-233.

Gardner, R. W., & Schoen, R. A. Differentiation and abstraction in concept formation. Psychological Monographs, 1962, 76, No. 41 (Whole No. 560).

Harvey, O. J. Scoring manual for the This I Believe Test. Boulder: University of Colorado, undated (mimeo).

Harvey, O. J. System structure, flexibility and creativity. In O. J. Harvey (Ed.), Experience, structure and adaptability. New York: Springer, 1966. Pp. 39-65.

Harvey, O. J., Hunt, D. E., & Schroder, H. M. Conceptual systems and personality organization. New York: Wiley, 1961.

Harvey, O. J., & Schroder, H. M. Cognitive aspects of self and motivation. In O. J. Harvey (Ed.), Motivation and social interaction. New York: Ronald Press, 1963. Pp. 95-133.

Hess, H. F. Additional evidence on the generality of cognitive complexity. Perceptual and Motor Skills, 1966, 23, 857-858.

Kelly, G. A. The psychology of personal constructs. New York: Norton, 1955, 2 vols.

Messick, S., & Kogan, N. Differentiation and compartmentalization in object-sorting measures of categorizing style. Perceptual and Motor Skills, 1963, 16, 47-51.

Pettigrew, T. F. The measurement and correlates of category width as a cognitive variable. Journal of Personality, 1958, 26, 532-544.

Rozeboom, W. W. Foundations of the theory of prediction. Homewood, Ill.: The Dorsey Press, 1966.

Schroder, H. M. Conceptual complexity and personality organization. Princeton, N. J.: Princeton University, 1969 (mimeo).

Schroder, H. M., Driver, M. J., & Streufert, S. Human information processing. New York: Holt, Rinehart & Winston, 1967.

Schroder, H. M., & Harvey, O. J. Conceptual organization and group structure. In O. J. Harvey (Ed.), Motivation and social interaction--cognitive determinants. New York: Ronald Press, 1963. Pp. 134-166.

Scott, W. A. Cognitive complexity and cognitive flexibility. Sociometry, 1962, 25, 405-414.

Scott, W. A. Brief report: Measures of cognitive structure. Multivariate Behavioral Research, 1966, 1, 391-395.

Scott, W. A. The structure of natural cognitions. Boulder: University of Colorado, 1967 (mimeo).

Scott, W. A. Structure of natural cognitions. Journal of Personality and Social Psychology, 1969, 12, 261-278.

Sloane, H. N., Gorlow, L., & Jackson, D. N. Cognitive styles in equivalence range. Perceptual and Motor Skills, 1963, 16, 389-404.

Todd, F. J., & Rappoport, L. A cognitive structure approach to person perception: A comparison of two models. Journal of Abnormal and Social Psychology, 1964, 68, 469-478.

Vannoy, J. S. A study of the generality of cognitive complexity as a personality construct. (Doctoral dissertation, University of Illinois) Ann Arbor, Mich.: University Microfilms, 1964. No. 65-3689.

Vannoy, J. S. Generality of cognitive complexity--Simplicity as a personality construct. Journal of Personality and Social Psychology, 1965, 2, 385-396.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wolfe, R. The role of conceptual systems in cognitive functioning at varying levels of age and intelligence. Journal of Personality, 1963, 31, 108-123.

Zajonc, R. B. The process of cognitive tuning in communication. Journal of Abnormal and Social Psychology, 1960, 61, 159-167.

BIOGRAPHY

Name:	Gary Bruce Cox.
Born:	November 6, 1939, in Bend, Oregon.
Married:	Carol Joan Hall, 1966.
Educational History:	Public Schools, Bend, Oregon, 1945-1957.
	Willamette University, Salem, Oregon, 1957-1959.
	University of Illinois, 1959-1960.
	University of Oregon, 1961-1962 (B.S.).
	Duke University, 1963-1969.
Professional Experience:	Research Assistant, Mental Health Research Institute, Ft. Steilacoom, Washington, 1962-1963.
	USPHS Traineeship, Duke University, 1963-64.
	VA Traineeship, 1964-65.
	USPHS Traineeship, 1965-66.
	USPHS Predoctoral Fellowship, 1966-67.
	Assistant Professor, Johnson C. Smith University, 1967-70.

Duke University Libraries
D00553035L

DUKE LSC
D00553035L